Advances in Experimental Medicine and Biology 945

Albert Jeltsch
Renata Z. Jurkowska *Editors*

DNA Methyltransferases – Role and Function

Advances in Experimental Medicine and Biology

Volume 945

Editorial Board
IRUN R. COHEN, *The Weizmann Institute of Science, Rehovot, Israel*
ABEL LAJTHA, *N.S. Kline Institute for Psychiatric Research, Orangeburg, NY, USA*
JOHN D. LAMBRIS, *University of Pennsylvania, Philadelphia, PA, USA*
RODOLFO PAOLETTI, *University of Milan, Milan, Italy*

Advances in Experimental Medicine and Biology presents multidisciplinary and dynamic findings in the broad fields of experimental medicine and biology. The wide variety in topics it presents offers readers multiple perspectives on a variety of disciplines including neuroscience, microbiology, immunology, biochemistry, biomedical engineering and cancer research. Advances in Experimental Medicine and Biology has been publishing exceptional works in the field for over 30 years and is indexed in Medline, Scopus, EMBASE, BIOSIS, Biological Abstracts, CSA, Biological Sciences and Living Resources (ASFA-1), and Biological Sciences. The series also provides scientists with up to date information on emerging topics and techniques.

More information about this series at http://www.springer.com/series/5584

Albert Jeltsch • Renata Z. Jurkowska
Editors

DNA Methyltransferases - Role and Function

Editors
Albert Jeltsch
Institut für Biochemie
Universität Stuttgart
Stuttgart
Germany

Renata Z. Jurkowska
BioMed X Innovation Center
Heidelberg
Germany

ISSN 0065-2598 ISSN 2214-8019 (electronic)
Advances in Experimental Medicine and Biology
ISBN 978-3-319-43622-7 ISBN 978-3-319-43624-1 (eBook)
DOI 10.1007/978-3-319-43624-1

Library of Congress Control Number: 2016957758

© Springer International Publishing Switzerland 2016
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland
The registered company address is Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

DNA is the key to the inheritance of the genetic information in living organisms, and the mechanism of duplication of double-stranded DNA is arguably the molecular process with the widest appreciation in the general public. However, almost since their discovery, it was known that the nucleobases could be further modified by the addition of methyl groups. DNA methylation patterns are often heritable, leading to their classification as epigenetic mark. We now know that DNA methylation has very important functions in almost all species, ranging from bacteria to higher and low eukaryotes and plants. Moreover, changes in DNA methylation are associated with the development of human diseases, and the field of epigenetics is currently exploding with connections to nutrition, behavior, and transgenerational inheritance of traits. The comprehension of the relevance of DNA methylation in various fields of biology and medicine has also brought a lot of attention to the enzymes responsible for the transfer of methyl group to DNA, the so-called DNA methyltransferases (MTases), which are the key subject of this book.

We both have studied DNA methyltransferases for many years, attracted by their complicated mechanisms, beautiful structures and medical relevance; actually, these enzymes faithfully accompanied us through almost our entire career. Therefore, we felt very honored and excited by the offer from *Springer* to edit a book on these fascinating enzymes and happily took up this challenge. This book on DNA methyltransferases provides a compilation of chapters that recapitulate and update many of the developments made in the field, including past achievements and future challenges. Many of the chapters were written by renowned experts, who themselves made central contributions to the developing field.

The introduction of the book (Chap. 1) by Jurkowska and Jeltsch recaptures the development of the field over the past more than 60 years, highlighting and conceptualizing many critical key discoveries. In Chap. 2, Motorin et al. place DNA methylation and DNA MTases into the larger subject of nucleic acid modification focusing on the alkylation of pyrimidines in RNA and DNA. Chapter 3 written by Casadesus describes bacterial DNA methyltransferases and the important roles of DNA methylation in bacteria. The next four chapters cover DNA methylation and DNA MTases in mammals. Tajima et al. focus in Chap. 4 on the structural aspects of the mammalian DNA MTases, and Jurkowska and Jeltsch describe their enzymatic properties and regulation in cells (Chap. 5). In Chap. 6, Dan and Chen review the important contributions of genetic studies to our current understanding

of DNA methylation and DNA MTases. Chapter 7 by Lakshminarasimhan and Liang recapitulates the role of DNA methylation in cancer. Next, structures and mechanisms of plant DNA methyltransferases are described in Chap. 8 written by Du, and in Chap. 9, Wedd and Maleszka present the role of DNA methylation in honeybees as an example for DNA methylation in lower eukaryotes. The ongoing progress of science in the field is illustrated by the recent discovery of adenine-N6 methylation in several higher eukaryotes and the discussion whether this mark, still controversial in some cases, functions as an epigenetic signal, as described in Chap. 10 by O'Brown and Greer. The next chapters focus on the pathways of DNA demethylation (Chap. 11 written by Dean) and the structure and mechanism TET enzymes, which are involved in this reaction (Chap. 12 contributed by Yin and Xu). In Chap. 13, Shimbo and Wade summarize the biological processes involved in DNA methylation readout, and Hong and Cheng review base flipping as a basic mechanism involved in setting, reading, and erasing DNA methylation in Chap. 14. The last part of the book is devoted to the technological developments. In Chap. 15, Tost reviews current methods to study DNA methylation. Based on this, Lopez et al. provide an overview over the development and potential application of DNMT inhibitors in cancer and other diseases (Chap. 16), and Stolzenburg et al. describe emerging approaches to edit DNA methylation patterns in a targeted manner (Chap. 17). Finally, in Chap. 18, Laurino et al. summarize achievements in the design of DNA methyltransferases, and in Chap. 19, Tomkuvienė et al. describe applications of DNMTs as molecular biology tools to label DNA.

We anticipate many more years of exciting research focusing on DNA methylation and DNA MTases, with many new and groundbreaking discoveries to come. The aim of this book is to serve as a rich and reliable source of information for specialist scientists, but also for students and researchers entering the field, providing them with a solid fundament for future work. At the same time, it should help researchers to get into this fascinating subject, allowing them to catch up with the current level of knowledge and learn about recent trends.

Stuttgart, Germany
May 2016

Renata Z. Jurkowska
Albert Jeltsch

Contents

Mechanisms and Biological Roles of DNA Methyltransferases and DNA Methylation: From Past Achievements to Future Challenges 1
Renata Z. Jurkowska and Albert Jeltsch

DNA and RNA Pyrimidine Nucleobase Alkylation at the Carbon-5 Position 19
Yuri Motorin, Salifu Seidu-Larry, and Mark Helm

Bacterial DNA Methylation and Methylomes 35
Josep Casadesús

Domain Structure of the Dnmt1, Dnmt3a, and Dnmt3b DNA Methyltransferases 63
Shoji Tajima, Isao Suetake, Kohei Takeshita, Atsushi Nakagawa, and Hironobu Kimura

Enzymology of Mammalian DNA Methyltransferases 87
Renata Z. Jurkowska and Albert Jeltsch

Genetic Studies on Mammalian DNA Methyltransferases 123
Jiameng Dan and Taiping Chen

The Role of DNA Methylation in Cancer 151
Ranjani Lakshminarasimhan and Gangning Liang

Structure and Mechanism of Plant DNA Methyltransferases 173
Jiamu Du

DNA Methylation and Gene Regulation in Honeybees: From Genome-Wide Analyses to Obligatory Epialleles 193
Laura Wedd and Ryszard Maleszka

N6-Methyladenine: A Conserved and Dynamic DNA Mark 213
Zach Klapholz O'Brown and Eric Lieberman Greer

Pathways of DNA Demethylation 247
Wendy Dean

Structure and Function of TET Enzymes 275
Xiaotong Yin and Yanhui Xu

Proteins That Read DNA Methylation 303
Takashi Shimbo and Paul A. Wade

DNA Base Flipping: A General Mechanism for Writing, Reading, and Erasing DNA Modifications 321
Samuel Hong and Xiaodong Cheng

Current and Emerging Technologies for the Analysis of the Genome-Wide and Locus-Specific DNA Methylation Patterns 343
Jörg Tost

DNA Methyltransferase Inhibitors: Development and Applications 431
Marie Lopez, Ludovic Halby, and Paola B. Arimondo

Rewriting DNA Methylation Signatures at Will: The *Curable Genome* Within Reach? 475
Sabine Stolzenburg, Desiree Goubert, and Marianne G. Rots

Engineering and Directed Evolution of DNA Methyltransferases 491
Paola Laurino, Liat Rockah-Shmuel, and Dan S. Tawfik

DNA Labeling Using DNA Methyltransferases 511
Miglė Tomkuvienė, Edita Kriukienė, and Saulius Klimašauskas

Mechanisms and Biological Roles of DNA Methyltransferases and DNA Methylation: From Past Achievements to Future Challenges

Renata Z. Jurkowska and Albert Jeltsch

Abstract

DNA methylation and DNA methyltransferases (MTases) – the enzymes that introduce the methylation mark into the DNA – have been studied for almost 70 years. In this chapter, we review key developments in the field that led to our current understanding of the structures and mechanisms of DNA MTases and the essential biological role of DNA methylation, including the discovery of DNA methylation and DNA MTases, the cloning and sequence analysis of bacterial and eukaryotic MTases, and the elucidation of their structure, mechanism, and regulation. We describe genetic studies that contributed greatly to the evolving views on the role of DNA methylation in human development and diseases, the invention of methods for the genome-wide analysis of DNA methylation, and the biochemical identification of DNA MTases and the family of TET enzymes, which are involved in DNA demethylation. We finish by highlighting critical questions for the next years of research in the field.

R.Z. Jurkowska (✉)
BioMed X Innovation Center, Im Neuenheimer Feld 583, Heidelberg D-69120, Germany
e-mail: jurkowska@bio.mx

A. Jeltsch (✉)
Faculty of Chemistry, Institute of Biochemistry, University of Stuttgart, Pfaffenwaldring 55, Stuttgart D-70569, Germany
e-mail: albert.jeltsch@ibc.uni-stuttgart.de

© Springer International Publishing Switzerland 2016

A. Jeltsch, R.Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*, Advances in Experimental Medicine and Biology 945,
DOI 10.1007/978-3-319-43624-1_1

Abbreviations

AdoMet	S-Adenosyl-L-methionine
HPLC	High-pressure liquid chromatography
MTase	DNA methyltransferase
RM system	Restriction/modification system
SMRT sequencing	Single-molecule real-time sequencing
TET enzyme	Ten-eleven translocation enzyme
TLC	Thin-layer chromatography

1 Discovery of DNA Methylation

DNA from various sources contains the methylated bases C5-methylcytosine, N4-methylcytosine, and N6-methyladenine in addition to the four standard nucleobases (Fig. 1a). Methylation of cytosine at the C5-position has been discovered in calf thymus DNA already in 1948 using paper chromatography experiments (Hotchkiss 1948), and 6-methyladenine was found in bacterial DNA in 1955 (Dunn and Smith 1955). N4-methylcytosine, the third and least common methylated base in bacterial DNA, was described for the first time in 1983 (Janulaitis et al. 1983). The methylation of nucleobases at these positions places the methyl groups in the major groove of double-stranded B-DNA, where they do not interfere with the Watson/Crick base pairing, but can easily be detected by proteins interacting with the DNA (Fig. 1b). By this mechanism, the methylation adds extra information to the DNA that is not encoded in the DNA sequence, and the methylated bases can be

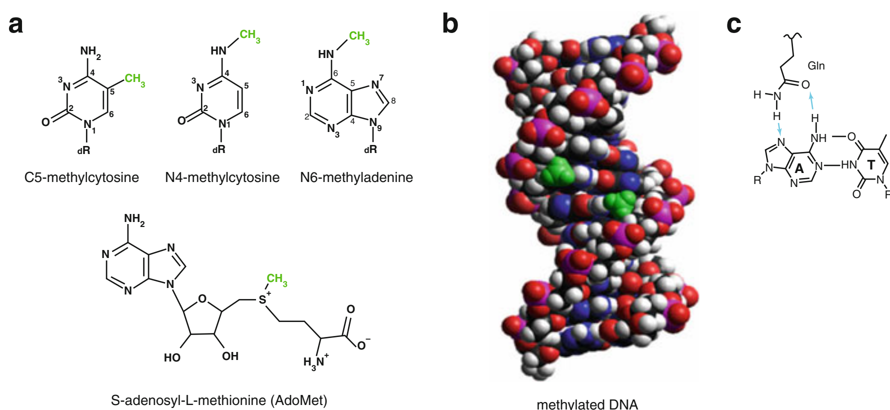


Fig. 1 Molecules related to DNA methylation. (**a**) Structures of the methylated bases that occur in DNA and of the AdoMet cofactor, the universal methyl group donor for all DNA methylation reactions. (**b**) Space-filling model of the structure of B-DNA with a methylated CpG site. The methyl groups are shown in green in the major groove of the DNA. (**c**) Example of the major groove readout of an AT base pair by Gln as proposed by Seeman and colleagues in 1976 (Seeman et al. 1976). This contact would be disrupted by methylation of the A at the N6-position

considered the 5th, 6th, and 7th letters of the genetic alphabet (Jeltsch 2002). DNA methylation can, for example, directly prevent the readout of an AT base pair by Gln in the major groove (Fig. 1b, c). By this and related processes, DNA methylation can control the binding of proteins to DNA and thereby the expression of the genetic information. Despite the interesting properties of methylated bases and their importance in living organisms, many years had passed after the initial discovery of DNA methylation until work with DNA methyltransferases, the enzymes that introduce this modification, had been systematically started.

2 Discovery and Early Work on DNA MTases

DNA methyltransferases were initially discovered as parts of restriction/modification (RM) systems (Arber and Dussoix 1962). S-Adenosyl-L-methionine (AdoMet)-dependent DNA and RNA methylation activity was first described by Gold in 1963 (Gold et al. 1963) and a series of papers published by Gold in 1964 (Gold et al. 1964; Gold and Hurwitz 1964a, b; Hurwitz et al. 1964a, b). The *E. coli* EcoDam (a solitary bacterial MTase that is not part of an RM system) was initially described in 1973 (Marinus and Morris 1973) and purified in 1982 (Herman and Modrich 1982). The first studies with human and murine enzymes were reported in the late 1970s and early 1980s (Browne et al. 1977; Gruenbaum et al. 1982). However, in the 1970s and 1980s, DNA MTases remained a kind of passengers in the ongoing molecular biology revolution, due to their functional and genetic association to restriction endonucleases, which were absolutely essential as analytical and cloning tools at this time (Arber and Linn 1969; Boyer 1971; Meselson et al. 1972). In addition, restriction endonucleases and DNA MTases constituted the first model system to study sequence-specific DNA recognition, a process essential to the control of gene expression in all forms of life (Modrich 1982).

With the increasing commercial importance of restriction endonucleases, biotech companies were interested to shift the production procedures away from purification of enzymes from the original bacterial strains toward recombinant expression of cloned enzymes. Therefore, cloning of restriction enzymes moved into the center of the scientific and economic interest. It was known that RM systems often reside on mobile genetic elements, where the genes coding for the methyltransferase and the endonuclease are located next to each other. Hence, cloning of a DNA fragment containing the methyltransferase gene often led to the cloning of the restriction enzyme gene on the same DNA insert. In a procedure called "Hungarian trick," the group of Venetianer realized that the special properties of DNA methyltransferases could be exploited to selectively clone genes encoding these enzymes (Szomolanyi et al. 1980). This approach was based on the fact that after expression of a DNA MTase in cells, the enzyme modified its own encoding DNA. Hence, after shotgun cloning of bacterial genomes, the DNA inserts were isolated and cleaved with a restriction enzyme of interest. The protected DNA likely coded for a methyltransferase, which methylated DNA within the target region sequence of the endonuclease and thereby prevented cleavage. After cloning of these protected inserts, it turned out that very often the gene for the restriction

enzyme was found on the same piece of DNA next to the methyltransferase gene. Almost 20 years later, a similar coupling of genotype and phenotype after expression of DNA methyltransferases was applied by Tawfik and colleagues to develop a novel approach for protein engineering, which was based on the expression of libraries of MTase mutants in water/oil emulsions (Tawfik and Griffiths 1998).

3 DNA MTases Contain Conserved Amino Acid Sequence Motifs

The wide application of the above-described and related cloning procedures led to the cloning of hundreds of restriction enzymes together with their corresponding methyltransferases. Therefore, the group of bacterial DNA methyltransferases provided a rich source of enzymes recognizing different DNA sequences for enzymatic, biochemical, and evolutionary studies, which has led to many important insights and breakthrough discoveries (Wilson and Murray 1991; Pingoud and Jeltsch 1997; Pingoud et al. 2014). Comparisons of the amino acid sequences of various DNA methyltransferases in the early days of multiple sequence alignments led to the discovery of ten amino acid motifs characteristic for cytosine-C5 methyltransferases (Posfai et al. 1989; Klimasauskas et al. 1989; Lauster et al. 1989) (Fig. 2). In 1988, Bestor cloned the first mammalian DNA methyltransferase that was found to share extensive sequence similarity with the bacterial cytosine-C5 methyltransferases in its C-terminal catalytic part (Bestor et al. 1988). It was discovered that bacterial adenine-N6 methyltransferases contained conserved amino acid motifs as well (Fig. 2) (Lauster et al. 1987; Guschlbauer 1988), and some of the MTase motifs were shown to be part of general signature motifs of all AdoMet-dependent methyltransferases, also including small molecule, protein, and RNA methyltransferases (Kagan and Clarke 1994; Ingrosso et al. 1989). Although statistical methods were insufficient at that time, these studies led to the identification of the key catalytic regions, both in adenine-N6 and cytosine-C5 methyltransferases. Many of the most conserved residues in both families of enzymes were shown to be directly involved in the catalytic process (Cheng 1995; Jeltsch 2002), and several amino acid motifs identified in the early alignment studies could later be connected to defined structural elements in the conserved methyltransferase fold (Malone et al. 1995) (Fig. 2).

4 Structure and Mechanism of DNA MTases

All DNA methyltransferases use AdoMet as a methyl group donor. Based on their mechanism, one can distinguish methyltransferases adding the methyl group to carbon or nitrogen atoms. The former group comprises cytosine-C5, the latter adenine-N6 and cytosine-N4 methyltransferases. All DNA MTases follow a ternary complex mechanism, where the catalytically competent complex consists of the enzyme, the DNA substrate, and the AdoMet cofactor. In some enzymes, binding of the DNA substrate and the AdoMet is ordered; in other examples it is random. Wu and Santi studied the catalytic mechanism of cytosine-C5 methyltransferases and discovered in

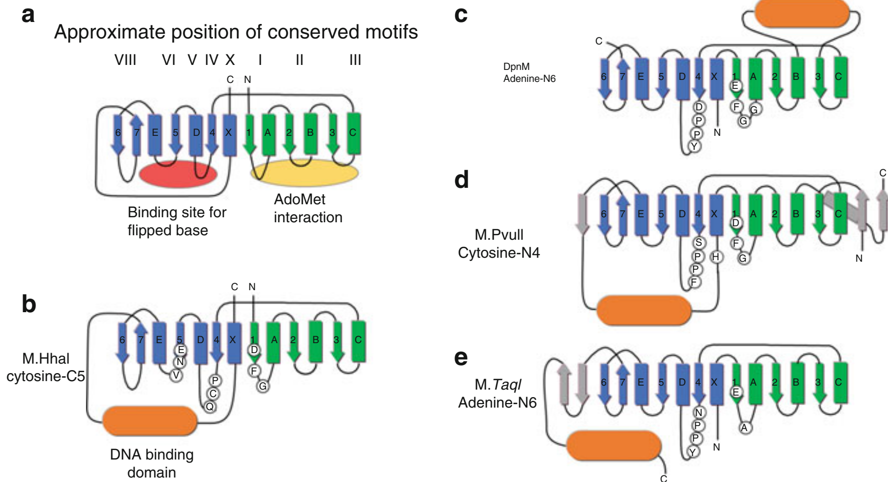


Fig. 2 Topological scheme of the universal AdoMet-dependent DNA MTase fold. It consists of two Rossmann fold half domains with several conserved amino acid sequences, one subdomain forming the binding site for the AdoMet and the second for the flipped base. DNA recognition is mediated by a DNA-binding domain, which is variable in sequence and structure. The linear arrangements of the functional elements vary between different DNA MTases by circular permutation, generating several characteristic subgroups of DNA MTases. (**a**) Schematic representation of the general structure of the DNA MTase fold. (**b**) General structure of cytosine-C5 MTases. (**c–e**) General structure of three subgroups of adenine-N6 and cytosine-N4 MTases

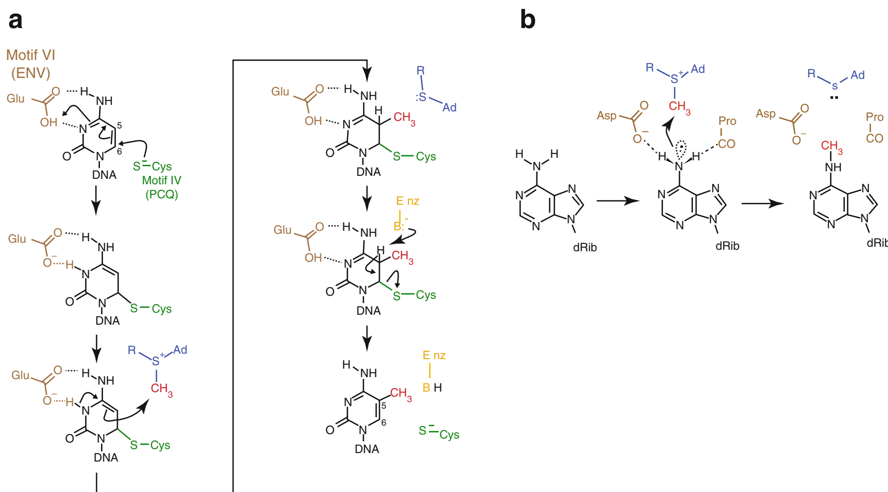


Fig. 3 Catalytic mechanism of DNA MTases. (**a**) Mechanism of DNA-(cytosine C5)-MTases. (**b**) Mechanism of DNA-(adenine N6)-MTases

1985 that it follows a Michael addition reaction, which is characterized by the formation of a covalent intermediate between the enzyme and the target base (Wu and Santi 1985, 1987) (Fig. 3a). Shortly afterward, Santi and coworkers also showed

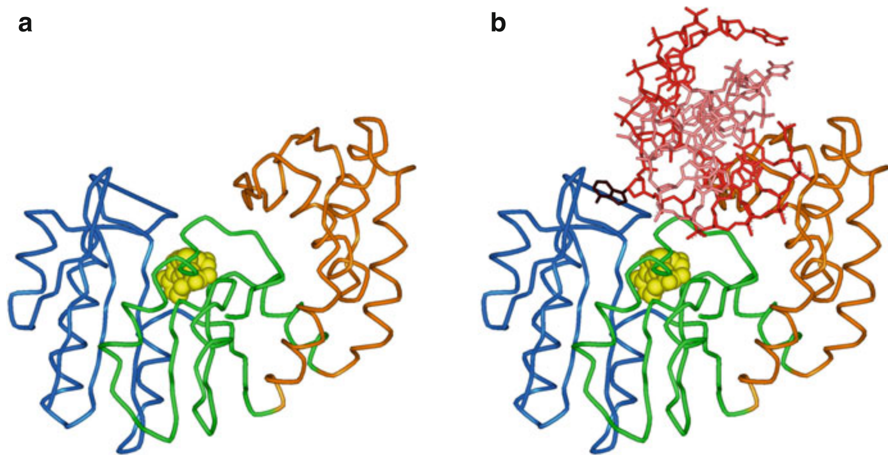


Fig. 4 Ribbon model of the structure of the EcoDam DNA MTase (Horton et al. 2006). The AdoMet and the base-binding subdomains are shown in *green* and *blue*, respectively. AdoMet is displayed in space-filling form in *yellow*. The DNA-binding domain is colored in *orange*. (**a**) Structure of the EcoDam-AdoMet complex. (**b**) Structure of the complex of EcoDam with bound substrate DNA (*red*, the flipped adenine base is shown in *black*)

that adenine methylation proceeds directly at the N6 position despite the poor nucleophilicity of the N6 atom and not by a transient transfer of the methyl group to the N1 followed by its shift to the N6 (Pogolotti et al. 1988). Seminal insights into the folding of the methyltransferases and the arrangement of their catalytic center came with the first structure of a DNA methyltransferase (the bacterial M.HhaI enzyme) that was solved in 1993 (Cheng et al. 1993). In 1994, the publication of the first structure of a DNA methyltransferase (again M.HhaI) with its DNA substrate by Cheng and coworkers led to another conceptual breakthrough regarding the catalytic mechanism of DNA methyltransferases (Klimasauskas et al. 1994) (Fig. 4). It was observed that the target base for the methylation reaction was completely rotated out of the DNA helix and inserted into a catalytic pocket (so-called base flipping), allowing access of the catalytic residues described above to the base. This unexpected and seminal discovery highlighted the flexibility of DNA and the dynamic processes that accompany enzymatic catalysis; research subjects that were intensively studied afterward. Today, we appreciate base flipping as a universal process in DNA methylation, but also in other reactions occurring on DNA, including DNA repair (Roberts 1995; Roberts and Cheng 1998). The first structure of an adenine-N6 MTase (M.TaqI) was published in 1994 as well (but without DNA) (Labahn et al. 1994), unexpectedly showing that both enzyme families contain a large catalytic domain with an identical fold, consisting of a six-stranded parallel ß-sheet with a seventh strand inserted in an antiparallel fashion between the fifth and sixth strands (Schluckebier et al. 1995) (Fig. 2). This fold is known today as the AdoMet-dependent methyltransferase fold (Martin and McMillan 2002). The seven-stranded ß-sheet is flanked by α-helices creating two subdomains with Rossmann fold

architecture: one containing the binding site for the AdoMet and the other for the flipped base (Cheng 1995; Jeltsch 2002). In addition, all MTases contain a second less-conserved domain involved in DNA recognition. The first structure of an adenine-N6 MTase with DNA was solved in 2001 (Goedecke et al. 2001). It showed that the N6 of the flipped adenine is positioned in a tetrahedral environment of hydrogen bond donors provided by the conserved residues of the (DNS)PP(YFW) motif, suggesting that its nucleophilicity is increased by a change in hybridization from sp2 to sp3 (Fig. 3b). Cytosine-N4 MTases are believed to follow an analogous mechanism, due to the chemical similarity of the methyl-acceptor atom and the observation that the specificity of enzymes form these families overlap, i.e. that adenine-N6 MTases can also methylate cytosine at N4 and cytosine-N4 MTases also methylate adenine (Jeltsch et al. 1999, 2001).

5 Molecular Evolution of MTases

As described above, the conserved structure of the 7-ß-strand MTases consists of two half domains with Rossmann folds, which are fused to each other. One of them mediates AdoMet interaction, and the second provides the binding sites for the methylation substrates, which are flipped nucleobases in the case of DNA MTases. The high structural similarity of all DNA MTases, and the presence of conserved motifs with similarities even between different groups of MTases, suggests that these enzymes are monophyletic. Presumably, the two subdomains originated from a duplication of a primordial AdoMet-binding Rossmann fold domain (Malone et al. 1995). Later, one subdomain continued to bind AdoMet, but the second diverged to generate the binding sites for different methylation substrates, including flipped cytosine and adenine bases, but also small molecules like catecholamine or amino acids like arginine, leading to the various groups of contemporary MTases specific for different methylation substrates. Moreover, the initial MTase ancestor has undergone several modifications during molecular evolution, including circular permutations (Jeltsch 1999; Bujnicki 2002), and in the case of DNA MTases, the insertion of diverse and unrelated DNA-binding domain at different places in the consensus structure, leading to the creation of different classes of DNA MTases (Malone et al. 1995) (Fig. 2).

6 Early Views on the Biological Role of DNA Methylation

Methylation of human and mammalian DNA at CpG sites was identified in the beginning of the 1980s (Razin and Riggs 1980; Ehrlich and Wang 1981); in plants DNA methylation was found also in CNG sites (where N is any nucleotide) (Gruenbaum et al. 1981) and at nonsymmetric sites. However, the early 1980s was a time when biology mainly focused on the detailed investigation of the so-called model organisms. While this approach was extremely farsighted and it greatly contributed to the explosion in our understanding of the molecular basis of life, it did not come without risk, as illustrated by the general lack of appreciation of DNA

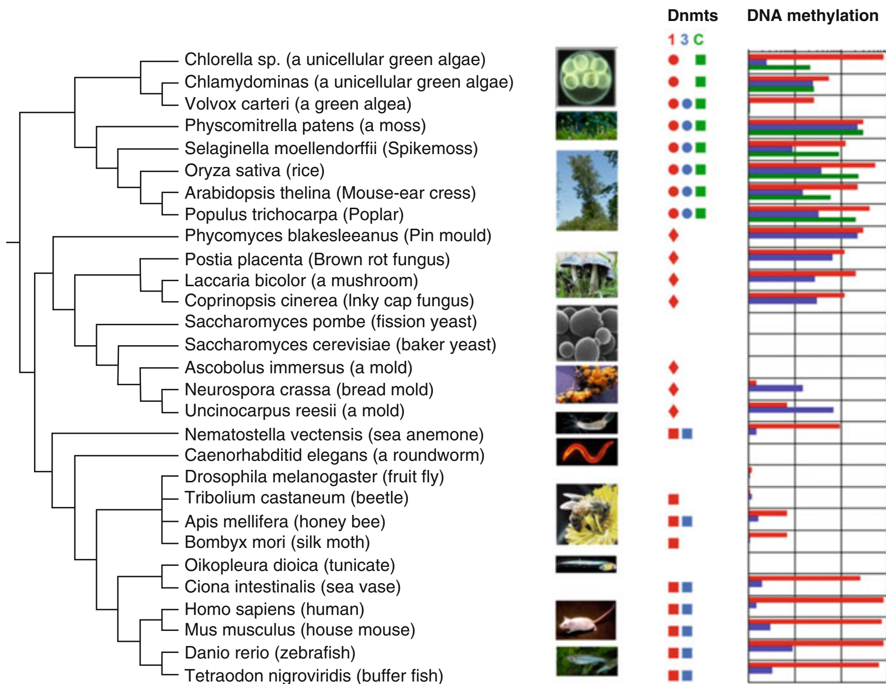


Fig. 5 Phylogenic distribution of DNA methylation systems and DNA MTases. The distribution of MTases of the DNMT1 (*red*), DNMT3 (*blue*), and chromomethylase families (*green*) is shown in several characteristic species. *Red circles* denote plant Met1 homologs, diamonds - enzymes of the fungal Dim-2 families and squares - DNMT1 homologs). *Blue circles* denote plant DRM homologs and squares (DNMT3 enzymes). DNA methylation data were averaged as described in Jeltsch (2010) and shown for CpG (*red*), CHH (*blue*), and CNG (*green*). CHG methylation is shown only for plants. The phylogenetic tree was generated with the National Center for Biotechnology Information taxonomy and the Interactive Tree of Life (Reproduced from Jeltsch (2010) with modifications. Reprinted with permission from AAAS)

methylation around this time. Due to an unfortunate coincidence, many of the carefully selected model organisms, like *S. cerevisiae*, *D. melanogaster*, or *C. elegans* were lacking detectable DNA methylation. Consistently, there was a widespread belief that DNA methylation, although having some interesting roles, cannot be very important. On the other hand, it became clear that DNA methylation had an enormous influence on the human genome, when Bird discovered the existence of the CpG islands (Bird 1980; Bird et al. 1985), which are defined as regions of high density of CpG sites within the genome that was already known to be globally depleted of this dinucleotide (Swartz et al. 1962). It was realized that the depletion of the CpG sites from the bulk genome was indirectly due to the mutagenic effect of cytosine-C5 methylation, leading to the preservation of CpG sites only if they were unmethylated, as in CpG islands. Today, we know that DNA methylation systems are found in almost all organisms and the model organisms listed above appear to be rather exceptions (Fig. 5).

7 Genetic Studies on DNMTs in Mammals

While models connecting DNA methylation with known epigenetic phenomena, gene expression, and development (see, e.g., (Riggs 1975; Holliday and Pugh 1975)) were developed, the general skepticism on the essential role of DNA methylation in human biology was only overcome with the discovery of the repressive function of DNA methylation on gene expression (Tazi and Bird 1990) and the finding that mice with a knockout of DNMT1 (Li et al. 1992), the only mammalian DNMT known by that time, die during early embryonic development in the uterus. However, as often in science, this discovery led to the next question, because it turned out that DNMT1 knockout cells were not completely devoid of DNA methylation (Lei et al. 1996), which opened a hunt for additional MTase enzymes. Researchers tried to purify additional DNA methyltransferases from human and mouse cells; in parallel, the rising flood of DNA sequences was searched for entries containing the characteristic DNA methyltransferase motifs described above. It was the bioinformatics approach that was successful at the end, leading to the discovery of the DNMT3A and DNMT3B enzymes in 1998 (Okano et al. 1998). Shortly afterward, both MTases were also shown to be essential in mice (Okano et al. 1999). Soon after, genetic studies showed that DNMT3A together with DNMT3L (a catalytically inactive paralog of DNMT3A and DNMT3B) were needed to set imping-ing marks in the mouse germline (Bourc'his et al. 2001; Bourc'his and Bestor 2004; Hata et al. 2002; Kaneda et al. 2004).

DNA methylation provides organisms with an efficient epigenetic regulation system, which is particularly important in multicellular organisms, which need to develop stable cellular differentiation. It has been speculated that the development of powerful epigenetic systems, comprising DNA methyltransferases, demethylases, and other enzyme systems introducing modifications on histones, has been a critical step in the evolution of multicellular life (Jeltsch 2013).

8 Structure, Function, and Regulation of Mammalian DNA MTases

More recently, structures of DNMT3A (Jia et al. 2007; Guo et al. 2015) and DNMT1 (Takeshita et al. 2011; Song et al. 2011, 2012) were published, showing that complicated regulatory processes, including oligomerization, conformational changes, and auto-inhibition, all interplay to accurately control the activity of these enzymes. In 1997, targeting of DNMT1 to replication foci via its interaction with PCNA was discovered (Chuang et al. 1997), but later it became clear that the interaction of DNMT1 with UHRF1 is essential for the targeting and DNMT1 activity as well (Bostick et al. 2007; Sharif et al. 2007). Furthermore, it was found that in addition to the indirect targeting by other complex partners, DNMTs directly interact with chromatin: DNMT3A and DNMT3B with their ADD domains binding to H3 tails unmethylated at K4 (Ooi et al. 2007; Zhang et al. 2010) and their PWWP domains binding to H3K36me3 (Dhayalan et al. 2010), and DNMT1 via its replication foci

targeting domain interacting with ubiquitinated H3 tails (Nishiyama et al. 2013). Moreover, the principles of the regulation of the activity and stability of DNMTs via posttranslational modifications begin to emerge (Esteve et al. 2011; Deplus et al. 2014), adding another fascinating layer to the study of these enzymes.

9 Discovery of TET Enzymes

A similar changeful journey as in the field of DNA de novo methylation was undertaken in the investigation of DNA demethylation, starting from the question of whether an active process of DNA demethylation might exist at all, leading to its discovery and the study of its mechanisms (Ooi and Bestor 2008). It was only in 2009, when the combination of powerful biochemical and bioinformatics approaches led to the discovery of the ten-eleven translocation (TET) enzymes (Tahiliani et al. 2009), which oxidize 5-methylcytosine to the hydroxymethyl, formyl, or carboxyl state, and the discovery of these modified bases in human DNA (Tahiliani et al. 2009; Kriaucionis and Heintz 2009; Munzel et al. 2010; Pfaffeneder et al. 2011; Ito et al. 2011; He et al. 2011). The exact role of these additional modified bases and the complete pathway of DNA demethylation are not yet fully understood (Wu and Zhang 2010; Hahn et al. 2014). Today, DNA methylation is recognized as an essential epigenetic mark that acts in concert with other chromatin modifications, like histone posttranslational modifications, histone variants, or noncoding RNA, and its genome-wide and locus-specific level is determined by the combined action of MTases, demethylases, and DNA replication (Jeltsch and Jurkowska 2014). In mammals, DNA methylation is involved in the epigenetic processes, like imprinting and X-chromosome inactivation, but it also has global roles in the generation of heterochromatin, silencing of repeats, and gene regulation during development and disease (Jurkowska et al. 2011).

10 Methods for Site-Specific Detection of DNA Methylation

The detection of DNA methylation for a long time was based on the initial methods, TLC (followed initially by HPLC and today by mass spectrometry), allowing for a quantitative overall genome methylation analysis, but without sequence resolution, and restriction digestion using enzymes sensitive to DNA methylation, allowing site-specific genome-wide analysis, but only at defined restriction sites. For cytosine-C5 methylation, this situation dramatically changed with the development of the bisulfite conversion method, which can be combined with a battery of downstream technologies to enable a genome-wide analysis of 5-methylcytosine at single-nucleotide resolution (Frommer et al. 1992; Clark et al. 1994). This technology in concert with the breakthroughs in DNA sequencing technology has enabled researchers starting in 2008 to provide first genome-wide DNA methylation maps of plant and mouse cells (Cokus et al. 2008; Lister et al. 2008; Meissner et al. 2008).

For N-methylation, it was only in 2010, almost 20 years after the discovery of the bisulfite technology, when the development of single-molecule real-time (SMRT)

sequencing for the first time provided a method for the genome-wide analysis of 6-methyladenine at single-nucleotide resolution (Flusberg et al. 2010). This invention was followed by a flurry of bacterial N6-adenine methylomes, including those of *E. coli* and *C. crescentus* (Sanchez-Romero et al. 2015), which provided novel insights into the role of DNA methylation in bacteria in defense mechanisms, cell division, gene expression, and DNA repair.

11 Role of DNA Methylation in Cancer

In 1983, first groups reported global hypomethylation of DNA in cancer cells (Feinberg and Vogelstein 1983; Gama-Sosa et al. 1983). Shortly afterwards, first examples of local hypermethylation of tumor suppressor gene promoters in cancer cells were discovered (Baylin et al. 1986; Greger et al. 1989). It is now well established that these two processes, global DNA hypomethylation and regional hypermethylation, occur in most tumor cells and are directly connected to the progression of the disease (Baylin 2012; Bergman and Cedar 2013). In 2010, it was been discovered that somatic mutations in DNMT3A are prevalently observed in AML patients; among them the R882H exchange was found with particularly high frequency (Yamashita et al. 2010). Later work has confirmed and extended this finding and showed that mutations in DNMT3A are drivers in the disease process (Hamidi et al. 2015).

The frequent observation of the inactivation of tumor suppressor genes in cancers by hypermethylation has prompted the development of DNA methyltransferase inhibitors for clinical applications. This field was pioneered by Jones with the development of 5-azacytidine (Jones and Taylor 1980), which afterwards was confirmed to form an irreversible covalent complex with DNA methyltransferases (Santi et al. 1984). Later, cofactor analogs were also introduced to inhibit DNA methyltransferases (Reich and Mashhoon 1990). Today, many more derivatives of these compounds were developed, and several of them are in clinical use for the treatment of cancer and other diseases (Yang et al. 2010; Fahy et al. 2012). In addition, in 1997, Xu and Bestor developed the targeted methylation approach, a method in which a DNA MTase is fused to a DNA-binding domain that targets the fusion protein to specific genomic loci and results in the introduction of DNA methylation at these sites (Xu and Bestor 1997). In combination with targeting of other epigenetic enzymes, this method has the capacity to alter the expression of disease-related genes by rewriting the epigenome and, thereby, might provide a causative cure to many diseases (Kungulovski and Jeltsch 2016).

12 Conclusions and Outlook

Although DNA MTases, the enzymes that introduce methylation into DNA, have been intensively studied, the interest in these enzymes is constantly increasing (Fig. 6). This is due to an ever-growing importance of DNA methylation as an epigenetic modification in organismic development and human diseases. Despite

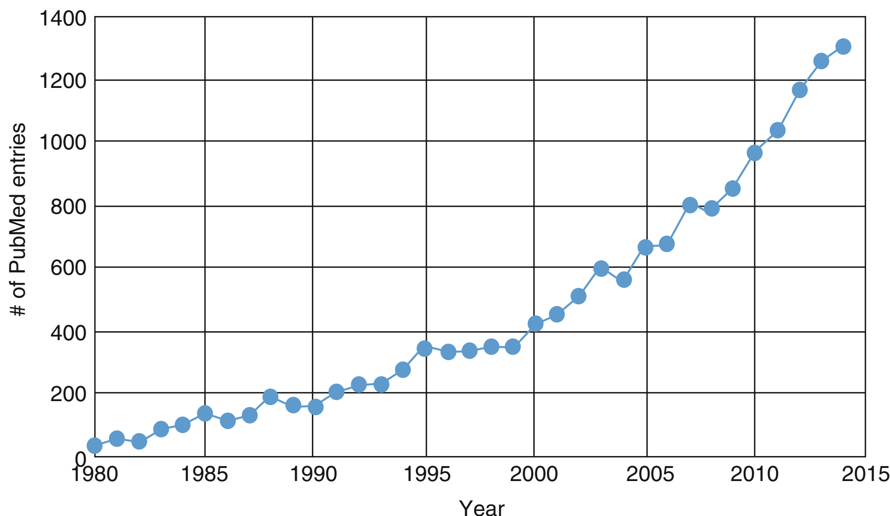


Fig. 6 Number of PubMed entries with the term "DNA" and "methyltransferase" in title or abstract (as of May, 2016)

decades of active research in the fields of DNA methylation and DNA methyltransferases, many exciting questions still await answers and future challenges extend from our current level of knowledge. How is DNA methylation (and epigenetic information in general) deposited during organismic development and how is it maintained and altered if needed? How are DNA MTases regulated and targeted to achieve these goals? How does DNA methylation interact with other epigenetic systems in mammals, lower eukaryotes, and even bacteria? Will it be soon possible to develop epigenetic antibacterial drugs, addressing processes, like phase variation or drug resistance? How can we make use of epigenetic editing, including targeted DNA methylation, to combat diseases like cancer? How does DNA demethylation work in detail? Which biological function(s) have the oxidized forms of 5-methylcytosine? Is there an active demethylation of N6-methyladenine in the cell? What is the role of this modification in other higher organisms? We anticipate many more years of exciting research to come in the field of DNA methylation, and the study of DNA MTases is an integral objective in this development.

References

Arber W, Dussoix D. Host specificity of DNA produced by Escherichia coli. I. Host controlled modification of bacteriophage lambda. J Mol Biol. 1962;5:18–36.

Arber W, Linn S. DNA modification and restriction. Annu Rev Biochem. 1969;38:467–500.

Baylin SB. The cancer epigenome: its origins, contributions to tumorigenesis, and translational implications. Proc Am Thorac Soc. 2012;9(2):64–5.

Baylin SB, Hoppener JW, de Bustros A, Steenbergh PH, Lips CJ, Nelkin BD. DNA methylation patterns of the calcitonin gene in human lung cancers and lymphomas. Cancer Res. 1986;46(6):2917–22.

Bergman Y, Cedar H. DNA methylation dynamics in health and disease. Nat Struct Mol Biol. 2013;20(3):274–81.
Bestor T, Laudano A, Mattaliano R, Ingram V. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. J Mol Biol. 1988;203(4):971–83.
Bird AP. DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res. 1980;8(7):1499–504.
Bird A, Taggart M, Frommer M, Miller OJ, Macleod D. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell. 1985;40(1):91–9.
Bostick M, Kim JK, Esteve PO, Clark A, Pradhan S, Jacobsen SE. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science. 2007;317(5845):1760–4.
Bourc'his D, Bestor TH. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature. 2004;431(7004):96–9.
Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH. Dnmt3L and the establishment of maternal genomic imprints. Science. 2001;294(5551):2536–9.
Boyer HW. DNA restriction and modification mechanisms in bacteria. Annu Rev Microbiol. 1971;25:153–76.
Browne MJ, Turnbull JF, McKay EL, Adams RL, Burdon RH. The sequence specificity of a mammalian DNA methylase. Nucleic Acids Res. 1977;4(4):1039–45.
Bujnicki JM. Sequence permutations in the molecular evolution of DNA methyltransferases. BMC Evol Biol. 2002;2:3.
Cheng X. Structure and function of DNA methyltransferases. Annu Rev Biophys Biomol Struct. 1995;24:293–318.
Cheng X, Kumar S, Posfai J, Pflugrath JW, Roberts RJ. Crystal structure of the HhaI DNA methyltransferase complexed with S-adenosyl-L-methionine. Cell. 1993;74(2):299–307.
Chuang LS, Ian HI, Koh TW, Ng HH, Xu G, Li BF. Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. Science. 1997;277(5334):1996–2000.
Clark SJ, Harrison J, Paul CL, Frommer M. High sensitivity mapping of methylated cytosines. Nucleic Acids Res. 1994;22(15):2990–7.
Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, et al. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature. 2008;452(7184):215–9.
Deplus R, Blanchon L, Rajavelu A, Boukaba A, Defrance M, Luciani J, et al. Regulation of DNA methylation patterns by CK2-mediated phosphorylation of Dnmt3a. Cell Rep. 2014;8(3):743–53.
Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S, et al. The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. J Biol Chem. 2010;285(34):26114–20.
Dunn DB, Smith JD. Occurrence of a new base in the deoxyribonucleic acid of a strain of Bacterium coli. Nature. 1955;175(4451):336–7.
Ehrlich M, Wang RY. 5-Methylcytosine in eukaryotic DNA. Science. 1981;212(4501):1350–7.
Esteve PO, Chang Y, Samaranayake M, Upadhyay AK, Horton JR, Feehery GR, et al. A methylation and phosphorylation switch between an adjacent lysine and serine determines human DNMT1 stability. Nat Struct Mol Biol. 2011;18(1):42–8.
Fahy J, Jeltsch A, Arimondo PB. DNA methyltransferase inhibitors in cancer: a chemical and therapeutic patent overview and selected clinical studies. Expert Opin Ther Pat. 2012;22(12):1427–42.
Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature. 1983;301(5895):89–92.
Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods. 2010; 7(6):461–5.
Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci U S A. 1992;89(5):1827–31.

Gama-Sosa MA, Slagel VA, Trewyn RW, Oxenhandler R, Kuo KC, Gehrke CW, et al. The 5-methylcytosine content of DNA from human tumors. Nucleic Acids Res. 1983;11(19):6883–94.
Goedecke K, Pignot M, Goody RS, Scheidig AJ, Weinhold E. Structure of the N6-adenine DNA methyltransferase M.TaqI in complex with DNA and a cofactor analog. Nat Struct Biol. 2001;8(2):121–5.
Gold M, Hurwitz J. The enzymatic methylation of ribonucleic acid and deoxyribonucleic acid. V. Purification and properties of the deoxyribonucleic acid-methylating activity of Escherichia coli. J Biol Chem. 1964a;239:3858–65.
Gold M, Hurwitz J. The enzymatic methylation of ribonucleic acid and deoxyribonucleic acid. Vi. Further studies on the properties of the deoxyribonucleic acid methylation reaction. J Biol Chem. 1964b;239:3866–74.
Gold M, Hurwitz J, Anders M. The enzymatic methylation of RNA and DNA, Ii. On the species specificity of the methylation enzymes. Proc Natl Acad Sci U S A. 1963;50(1):164–9.
Gold M, Hausmann R, Maitra U, Hurwitz J. The enzymatic methylation of Rna and DNA 8 effects of bacteriophage infection on the activity of the methylating enzymes. Proc Natl Acad Sci U S A. 1964;52:292–7.
Greger V, Passarge E, Hopping W, Messmer E, Horsthemke B. Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. Hum Genet. 1989;83(2):155–8.
Gruenbaum Y, Naveh-Many T, Cedar H, Razin A. Sequence specificity of methylation in higher plant DNA. Nature. 1981;292(5826):860–2.
Gruenbaum Y, Cedar H, Razin A. Substrate and sequence specificity of a eukaryotic DNA methylase. Nature. 1982;295(5850):620–2.
Guo X, Wang L, Li J, Ding Z, Xiao J, Yin X, et al. Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature. 2015;517(7536):640–4.
Guschlbauer W. The DNA, and S-adenosylmethionine-binding regions of EcoDam and related methyltransferases. Gene. 1988;74(1):211–4.
Hahn MA, Szabo PE, Pfeifer GP. 5-Hydroxymethylcytosine: a stable or transient DNA modification? Genomics. 2014;104(5):314–23.
Hamidi T, Singh AK, Chen T. Genetic alterations of DNA methylation machinery in human diseases. Epigenomics. 2015;7(2):247–65.
Hata K, Okano M, Lei H, Li E. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Development. 2002;129(8):1983–93.
He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science. 2011;333(6047):1303–7.
Herman GE, Modrich P. Escherichia coli dam methylase. Physical and catalytic properties of the homogeneous enzyme. J Biol Chem. 1982;257(5):2605–12.
Holliday R, Pugh JE. DNA modification mechanisms and gene activity during development. Science. 1975;187(4173):226–32.
Horton JR, Liebert K, Bekes M, Jeltsch A, Cheng X. Structure and substrate recognition of the Escherichia coli DNA adenine methyltransferase. J Mol Biol. 2006;358(2):559–70.
Hotchkiss RD. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. J Biol Chem. 1948;175(1):315–32.
Hurwitz J, Gold M, Anders M. The enzymatic methylation of ribonucleic acid and deoxyribonucleic acid. 3. Purification of soluble ribonucleic acid-methylating enzymes. J Biol Chem. 1964a;239:3462–73.
Hurwitz J, Gold M, Anders M. The enzymatic methylation of ribonucleic acid and deoxyribonucleic acid. Iv. The properties of the soluble ribonucleic acid-methylating enzymes. J Biol Chem. 1964b;239:3474–82.
Ingrosso D, Fowler AV, Bleibaum J, Clarke S. Sequence of the D-aspartyl/L-isoaspartyl protein methyltransferase from human erythrocytes. Common sequence motifs for protein, DNA, RNA, and small molecule S-adenosylmethionine-dependent methyltransferases. J Biol Chem. 1989;264(33):20131–9.
Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science. 2011;333(6047):1300–3.

Janulaitis A, Klimasauskas S, Petrusyte M, Butkus V. Cytosine modification in DNA by BcnI methylase yields N4-methylcytosine. FEBS Lett. 1983;161(1):131–4.

Jeltsch A. Circular permutations in the molecular evolution of DNA methyltransferases. J Mol Evol. 1999;49(1):161–4.

Jeltsch A. The cytosine N4-methyltransferase M.PvuII also modifies adenine residues. Biol Chem. 2001;382(4):707–10.

Jeltsch A. Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. Chembiochem. 2002;3(4):274–93.

Jeltsch A. Molecular biology. Phylogeny Methylomes Sci. 2010;328(5980):837–8.

Jeltsch A. Oxygen, epigenetic signaling, and the evolution of early life. Trends Biochem Sci. 2013;38(4):172–6.

Jeltsch A, Jurkowska RZ. New concepts in DNA methylation. Trends Biochem Sci. 2014;39(7):310–8.

Jeltsch A, Christ F, Fatemi M, Roth M. On the substrate specificity of DNA methyltransferases. adenine-N6 DNA methyltransferases also modify cytosine residues at position N4. J Biol Chem. 1999;274(28):19538–44.

Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature. 2007;449(7159):248–51.

Jones PA, Taylor SM. Cellular differentiation, cytidine analogs and DNA methylation. Cell. 1980;20(1):85–93.

Jurkowska RZ, Jurkowski TP, Jeltsch A. Structure and function of mammalian DNA methyltransferases. Chembiochem. 2011;12(2):206–22.

Kagan RM, Clarke S. Widespread occurrence of three sequence motifs in diverse S-adenosylmethionine-dependent methyltransferases suggests a common structure for these enzymes. Arch Biochem Biophys. 1994;310(2):417–27.

Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, Li E, et al. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature. 2004;429(6994):900–3.

Klimasauskas S, Timinskas A, Menkevicius S, Butkiene D, Butkus V, Janulaitis A. Sequence motifs characteristic of DNA[cytosine-N4]methyltransferases: similarity to adenine and cytosine-C5 DNA-methylases. Nucleic Acids Res. 1989;17(23):9823–32.

Klimasauskas S, Kumar S, Roberts RJ, Cheng X. HhaI methyltransferase flips its target base out of the DNA helix. Cell. 1994;76(2):357–69.

Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science. 2009;324(5929):929–30.

Kungulovski G, Jeltsch A. Epigenome editing: state of the art, concepts, and perspectives. Trends Genet. 2016;32(2):101–13.

Labahn J, Granzin J, Schluckebier G, Robinson DP, Jack WE, Schildkraut I, et al. Three-dimensional structure of the adenine-specific DNA methyltransferase M.Taq I in complex with the cofactor S-adenosylmethionine. Proc Natl Acad Sci U S A. 1994;91(23):10957–61.

Lauster R, Kriebardis A, Guschlbauer W. The GATATC-modification enzyme EcoRV is closely related to the GATC-recognizing methyltransferases DpnII and dam from E. coli and phage T4. FEBS Lett. 1987;220(1):167–76.

Lauster R, Trautner TA, Noyer-Weidner M. Cytosine-specific type II DNA methyltransferases. A conserved enzyme core with variable target-recognizing domains. J Mol Biol. 1989;206(2):305–12.

Lei H, Oh SP, Okano M, Juttermann R, Goss KA, Jaenisch R, et al. De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development. 1996;122(10):3195–205.

Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell. 1992;69(6):915–26.

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, et al. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell. 2008;133(3):523–36.

Malone T, Blumenthal RM, Cheng X. Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. J Mol Biol. 1995;253(4):618–32.

Marinus MG, Morris NR. Isolation of deoxyribonucleic acid methylase mutants of Escherichia coli K-12. J Bacteriol. 1973;114(3):1143–50.
Martin JL, McMillan FM. SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. Curr Opin Struct Biol. 2002;12(6):783–93.
Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature. 2008;454(7205):766–70.
Meselson M, Yuan R, Heywood J. Restriction and modification of DNA. Annu Rev Biochem. 1972;41:447–66.
Modrich P. Studies on sequence recognition by type II restriction and modification enzymes. CRC Crit Rev Biochem. 1982;13(3):287–323.
Munzel M, Globisch D, Bruckl T, Wagner M, Welzmiller V, Michalakis S, et al. Quantification of the sixth DNA base hydroxymethylcytosine in the brain. Angew Chem Int Ed Engl. 2010;49(31):5375–7.
Nishiyama A, Yamaguchi L, Sharif J, Johmura Y, Kawamura T, Nakanishi K, et al. Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. Nature. 2013;502(7470):249–53.
Okano M, Xie S, Li E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet. 1998;19(3):219–20.
Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell. 1999;99(3):247–57.
Ooi SK, Bestor TH. The colorful history of active DNA demethylation. Cell. 2008;133(7):1145–8.
Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature. 2007;448(7154):714–7.
Pfaffeneder T, Hackner B, Truss M, Munzel M, Muller M, Deiml CA, et al. The discovery of 5-formylcytosine in embryonic stem cell DNA. Angew Chem Int Ed Engl. 2011;50(31):7008–12.
Pingoud A, Jeltsch A. Recognition and cleavage of DNA by type-II restriction endonucleases. Eur J Biochem/FEBS. 1997;246(1):1–22.
Pingoud A, Wilson GG, Wende W. Type II restriction endonucleases--a historical perspective and more. Nucleic Acids Res. 2014;42(12):7489–527.
Pogolotti Jr AL, Ono A, Subramaniam R, Santi DV. On the mechanism of DNA-adenine methylase. J Biol Chem. 1988;263(16):7461–4.
Posfai J, Bhagwat AS, Posfai G, Roberts RJ. Predictive motifs derived from cytosine methyltransferases. Nucleic Acids Res. 1989;17(7):2421–35.
Razin A, Riggs AD. DNA methylation and gene function. Science. 1980;210(4470):604–10.
Reich NO, Mashhoon N. Inhibition of EcoRI DNA methylase with cofactor analogs. J Biol Chem. 1990;265(15):8966–70.
Riggs AD. X inactivation, differentiation, and DNA methylation. Cytogenet Cell Genet. 1975;14(1):9–25.
Roberts RJ. On base flipping. Cell. 1995;82(1):9–12.
Roberts RJ, Cheng X. Base flipping. Annu Rev Biochem. 1998;67:181–98.
Sanchez-Romero MA, Cota I, Casadesus J. DNA methylation in bacteria: from the methyl group to the methylome. Curr Opin Microbiol. 2015;25:9–16.
Santi DV, Norment A, Garrett CE. Covalent bond formation between a DNA-cytosine methyltransferase and DNA containing 5-azacytosine. Proc Natl Acad Sci U S A. 1984;81(22):6993–7.
Schluckebier G, O'Gara M, Saenger W, Cheng X. Universal catalytic domain structure of AdoMet-dependent methyltransferases. J Mol Biol. 1995;247(1):16–20.
Seeman NC, Rosenberg JM, Rich A. Sequence-specific recognition of double helical nucleic acids by proteins. Proc Natl Acad Sci U S A. 1976;73(3):804–8.
Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, Endo TA, et al. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature. 2007;450(7171):908–12.
Song J, Rechkoblit O, Bestor TH, Patel DJ. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. Science. 2011;331(6020):1036–40.

Song J, Teplova M, Ishibe-Murakami S, Patel DJ. Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. Science. 2012;335(6069):709–12.

Swartz MN, Trautner TA, Kornberg A. Enzymatic synthesis of deoxyribonucleic acid. XI. Further studies on nearest neighbor base sequences in deoxyribonucleic acids. J Biol Chem. 1962;237:1961–7.

Szomolanyi E, Kiss A, Venetianer P. Cloning the modification methylase gene of Bacillus sphaericus R in Escherichia coli. Gene. 1980;10(3):219–25.

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2009;324(5929):930–5.

Takeshita K, Suetake I, Yamashita E, Suga M, Narita H, Nakagawa A, et al. Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). Proc Natl Acad Sci U S A. 2011;108(22):9055–9.

Tawfik DS, Griffiths AD. Man-made cell-like compartments for molecular evolution. Nat Biotechnol. 1998;16(7):652–6.

Tazi J, Bird A. Alternative chromatin structure at CpG islands. Cell. 1990;60(6):909–20.

Wilson GG, Murray NE. Restriction and modification systems. Annu Rev Genet. 1991;25:585–627.

Wu JC, Santi DV. On the mechanism and inhibition of DNA cytosine methyltransferases. Prog Clin Biol Res. 1985;198:119–29.

Wu JC, Santi DV. Kinetic and catalytic mechanism of HhaI methyltransferase. J Biol Chem. 1987;262(10):4778–86.

Wu SC, Zhang Y. Active DNA demethylation: many roads lead to Rome. Nat Rev Mol Cell Biol. 2010;11(9):607–20.

Xu GL, Bestor TH. Cytosine methylation targeted to pre-determined sequences. Nat Genet. 1997;17(4):376–8.

Yamashita Y, Yuan J, Suetake I, Suzuki H, Ishikawa Y, Choi YL, et al. Array-based genomic resequencing of human leukemia. Oncogene. 2010;29(25):3723–31.

Yang X, Lay F, Han H, Jones PA. Targeting DNA methylation for epigenetic therapy. Trends Pharmacol Sci. 2010;31(11):536–46.

Zhang Y, Jurkowska R, Soeroes S, Rajavelu A, Dhayalan A, Bock I, et al. Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. Nucleic Acids Res. 2010;38(13):4246–53.

DNA and RNA Pyrimidine Nucleobase Alkylation at the Carbon-5 Position

Yuri Motorin, Salifu Seidu-Larry, and Mark Helm

Abstract

The carbon 5 of pyrimidine nucleobases is a privileged position in terms of nucleoside modification in both DNA and RNA. The simplest modification of uridine at this position is methylation leading to thymine. Thymine is an integral part of the standard nucleobase repertoire of DNA that is synthesized at the nucleotide level. However, it also occurs in RNA, where it is synthesized posttranscriptionally at the polynucleotide level. The cytidine analogue 5-methylcytidine also occurs in both DNA and RNA, but is introduced at the polynucleotide level in both cases. The same applies to a plethora of additional derivatives found in nature, resulting either from a direct modification of the 5-position by electrophiles or by further derivatization of the 5-methylpyrimidines. Here, we review the structural diversity of these modified bases, the variety of cofactors that serve as carbon donors, and the common principles shared by enzymatic mechanisms generating them.

Y. Motorin (✉)
IMoPA UMR7365 CNRS-UL, BioPole de l'Université de Lorraine,
9 avenue de la Foret de Haye, 54505 Vandoeuvre-les-Nancy, France
e-mail: motorine5@univ-lorraine.fr

S. Seidu-Larry
Department of Biochemistry, University of Cape Coast, College of Agriculture and Natural Sciences, School of Biological Sciences, Cape Coast, Ghana

Institute of Pharmacy and Biochemistry, Johannes Gutenberg University Mainz,
Staudingerweg 5, 55128 Mainz, Germany

M. Helm (✉)
Institute of Pharmacy and Biochemistry, Johannes Gutenberg University Mainz,
Staudingerweg 5, 55128 Mainz, Germany
e-mail: mhelm@uni-mainz.de

© Springer International Publishing Switzerland 2016

A. Jeltsch, R.Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*,
Advances in Experimental Medicine and Biology 945,
DOI 10.1007/978-3-319-43624-1_2

1 Introduction

One of only two chemical differences between RNA and DNA is the presence of a methyl group in deoxythymidine (T, also abbreviated dT, dm^5U). This substituent at the carbon 5 is thus part of one out of four integral building blocks of DNA, while no methyl group is present in its counterpart ribouridine (U, also abbreviated rU). The fact that both are metabolically derived from uridine monophosphate (containing ribose) is one of several arguments often used to support the claim that DNA has evolved from RNA (Kun et al. 2015; Muller 2006). It is also a clear indication of the importance of this methyl group, which, from this perspective, constitutes a nucleoside modification. Remarkably, and in contrast to most other nucleoside modifications, the obligatory presence of that methyl group in DNA is cemented by its introduction prior to nucleotide polymerization. Interestingly, thymidine also occurs in RNA (ribothymidine, also rT, m^5U, rm^5U), most prominently as the namesake of the T(Tψ)-loop in transfer RNA, where it is introduced posttranscriptionally. Methylated versions of the sister nucleobase cytidine are also found in both DNA (5mC) and RNA (m^5C); however, all are introduced at the polynucleotide level.

Beyond simple methyl groups, numerous chemically more complex modifications of pyrimidines are known that contain a carbon modification at the 5-position. This group may easily represent the largest group of all known modifications of nucleobases, because it includes a large part of the modifications of the U34-position in the anticodon of transfer RNAs (tRNAs) with its several dozen species (Machnicka et al. 2013). Again, the vast majority of these modifications are introduced at the polynucleotide level, the only exceptions being 5-hydroxymethylpyrimidines (5hmC, 5hmU) and their glycosylated derivatives found in phage DNA, which are generated at the mononucleotide level (Gommers-Ampt and Borst 1995). Interestingly, the resulting triphosphate nucleotides are then incorporated by phage DNA polymerase despite being sterically encumbered to a large degree. Even more interestingly, this tolerance for modifications at the 5-position appears to be a general feature of nucleotide polymerases on both the DNA and RNA level. Most prominently, T7-RNA polymerase, arguably the most commonly used enzyme for RNA synthesis, efficiently incorporates triphosphate ribonucleotides of rT and rm^5C, and even more sterically, demanding carbon-5 modifications have been incorporated into RNA this way (Vaught et al. 2004). Similarly, synthetic modifications are available for incorporation into DNA in PCR reactions by non-phage polymerases (Vaught et al. 2010).

The enzymatic mechanisms of pyrimidine C-5 modification involved here are of particular interest, since the central step involves a C–C bond formation. This reaction type is of increased interest to organic chemists and in natural product metabolism. In a large number of cases, proper understanding of the mechanism of bond-forming reactions requires the identification of the nucleophilic partner on one hand and the electrophilic partner on the other hand. This is typically easy for the formation of C–N or C–O bonds but more sophisticated for C–C bonds. In the case at hand, the carbon-5 position in pyrimidines is catalytically activated to form an intermittent carbon nucleophile, while the carbon side chains result from an

electrophilic metabolite, such as S-adenosyl-L-methionine (AdoMet) or N^5,N^{10}-methylenetetrahydrofolate (CH_2-THF). Hence, before addressing the biocatalysis of pyrimidine alkylation, we will discuss the reactivity of the 5-position from the perspective of the organic chemist and then do the same with the various electrophilic carbon scaffolds supplied as cofactors by the modification enzymes. Only then will we discuss a series of enzymatic reactions, of 5-pyrimidine alkylation and related relevant processes. Of note, the mechanism of the methyl group oxidations by TET enzymes leading, e.g., to 5hmC and 5hmU (Fu et al. 2014; Pfaffeneder et al. 2014) will not be discussed here. Instead, we will turn to equally fascinating modifications typically found in transfer RNA (tRNA) featuring a uridine at 34-position of the anticodon loop. Although this group of modifications includes a bewildering variety of sophisticated chemical structures, the initial modification reaction bears similarities with the aforementioned relatively simple modifications.

1.1 Chemical Structure and Occurrence of Pyrimidine C-5 Modifications

A surprising variety of pyrimidine modifications at the 5-position are known as of today and have been known for some time. Permutation of functional groups at three positions on the pyrimidine nucleoside, namely, H vs. OH at the 2′-position, NH_2 vs. OH at the 4-position, and H vs. CH_3 at the 5-position, results in eight species of pyrimidine nucleosides. Ribothymidine (rT, Fig. 1a) is ubiquitous in tRNA and very frequent in ribosomal RNA (rRNA), but not known elsewhere (Motorin and Helm 2011). 5mC is present in bacterial DNA as a guard against restriction nucleases (Roberts et al. 2005), while its presence in promoter and coding regions of eukaryotic genes participates in the regulation of transcription (Bogdanovic and Gomez-Skarmeta 2014). The distribution of m^5C in RNA, which is yet more complicated, was recently reviewed by us (Motorin et al. 2010) and has since encountered renewed interest through transcriptome-wide studies (Burgess et al. 2015; Militello et al. 2014; Squires et al. 2012).

The number of chemical species dramatically increases if 5-modifications other than methyl groups are admitted into this perusal. For example, recent research has discovered, or rediscovered, oxidation products of the 5-methyl group, including 5-hydroxymethyl and 5-formyl derivatives (Fig. 1b) (Kriaucionis and Heintz 2009; Tahiliani et al. 2009; Pfaffeneder et al. 2011). While it is common knowledge that DNA obligatorily contains T as a 5-methylated pyrimidine nucleobase, a less well-known exception is that the abovementioned glycosylated derivatives of 5hmC and 5hmU (termed “J-base”) are not just spurious modifications in the DNA of certain phages, but exist as near quantitative surrogates of the conventional C and T nucleosides (Fig. 1b) (Gommers-Ampt and Borst 1995). The unglycosylated precursor 5hmC was discovered in phage DNA as early as 1953 (Wyatt and Cohen 1953).

In eukaryotes, the existence of 5-hydroxymethylcytidine in DNA is a more recent discovery (Kriaucionis and Heintz 2009; Tahiliani et al. 2009) with a strong impact in fields such as epigenetics and developmental biology, while the corresponding

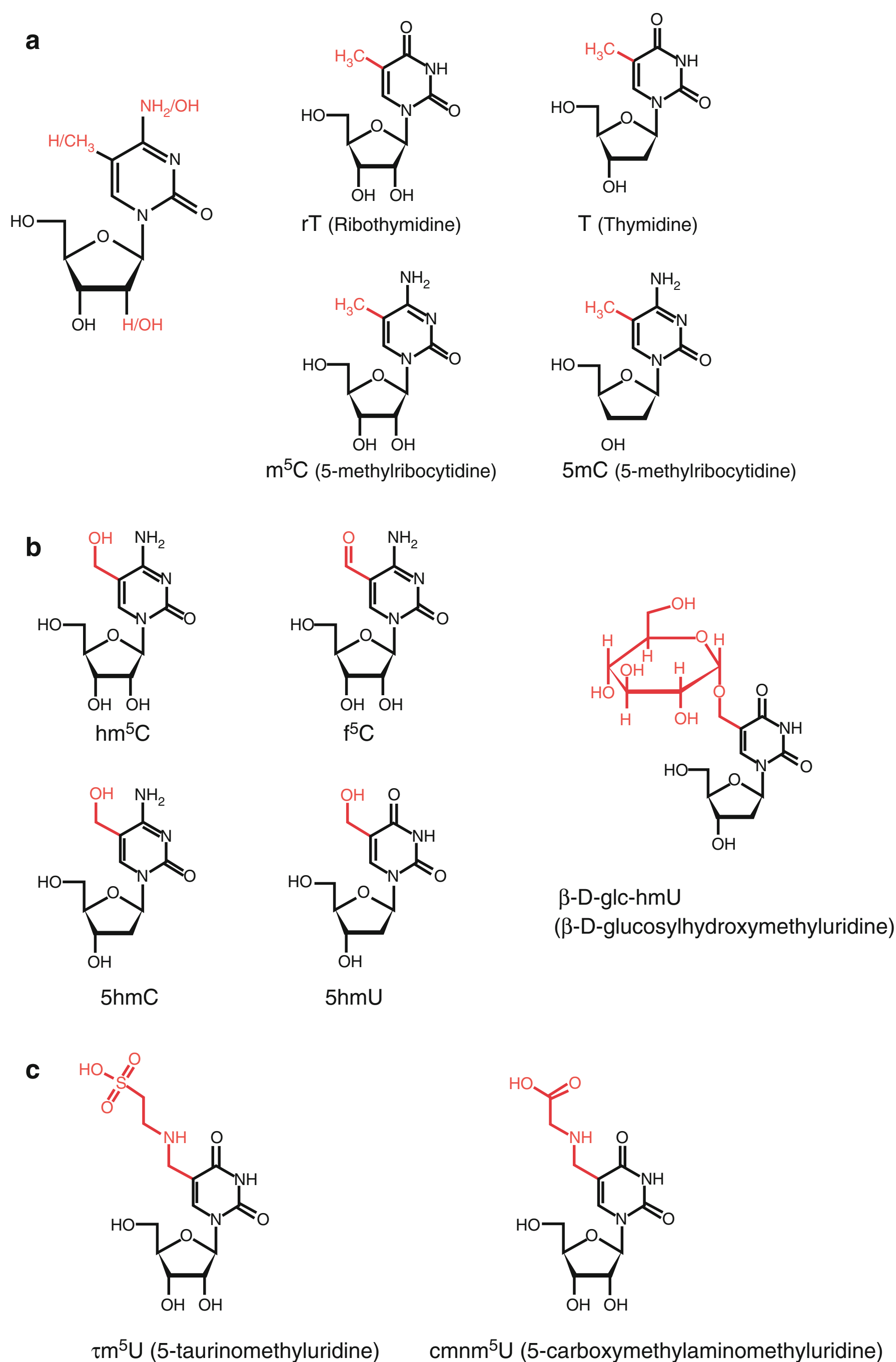


Fig. 1 DNA and RNA modifications arising from alkylation of the pyrimidine 5-position. (**a**) Thymidine, ribothymidine, m5C, and 5mC. (**b**) Modified nucleosides resulting from 5-mPy oxidation. (**c**) Modified uridines involved in codon recognition at 34-position of the tRNA anticodon

modification in RNA has also been reported decades ago, although incompletely characterized (Racz et al. 1978). Similarly, continued investigations have revealed 5-formylcytosine in DNA (fC, 5fC) (Pfaffeneder et al. 2011), while the corresponding ribonucleotide (f^5C) had been described in tRNA as early as 1994 (Moriya et al. 1994). However, further oxidation of 5fC leads to 5-carboxydeoxycytidine (caC, 5caC) (He et al. 2011), of which the ribonucleoside has yet to be discovered. 5hmU as a constituent of mammalian DNA has been discovered in traces and demonstrated to be a consequence of thymidine oxidation by TET enzymes (Pfaffeneder et al. 2014). Finally, the largest structural variety is found in aminomethyluridines, which are ribothymidine derivatives at the oxidation step of 5-hydroxymethyluridine and which predominate at 34-position in the anticodon of tRNAs (Machnicka et al. 2013). In contrast to 5hmU, these are not biochemically formed by oxidation of thymidine, but their biogenesis involves the use of a single carbon building block at the oxidation state of formaldehyde, namely, CH_2-THF, as will be detailed below.

1.2 Reactivity of Pyrimidines

A closer look at the catalytic strategies employed by modification enzymes acting on the 5-position of pyrimidines reveals that these exploit the intrinsic chemical reactivity of the pyrimidine ring. While this is not a surprising finding in general, the situation of pyrimidines is counterintuitive to the untrained biochemist, and a brief look at pyrimidine reactivity is conductive to a more intuitive mechanistic understanding of the involved enzyme.

Both nitrogen atoms within pyrimidines exert an electron withdrawing effect, resulting in an electron poor aromatic ring that is susceptible to nucleophiles. A nucleophilic attack, e.g., by bisulphite, at 6-position can be viewed as a Michael addition, while 4-position corresponds directly to the electrophilic center of a carbonyl functionality. Certain reactions with nucleophiles, such as hydrazine treatment, or the deamination reaction used in the so-called bisulphite sequencing (Schaefer et al. 2009; Frommer et al. 1992), sequentially exploit the electrophilic nature of both positions (Fig. 2).

In contrast, their electron poverty leaves pyrimidines relatively inert toward electrophilic reagents such as alkylating reagents, with the 3- and 5-positions being the exceptions (Motorin et al. 2010). The N3-position reacts with electrophilic reagents such as kethoxal or CMC, which is exploited in structural probing experiments (Giege et al. 1999). Concerning carbon 5, uridine reacts with the electrophilic formaldehyde under relatively mild acidic conditions to form 5-hydroxymethyluridine. Uracil (Kong et al. 2009), deoxyuridine (Conte et al. 1992), as well as cytidines (Khursid et al. 1982) are reported to also yield 5-hydroxymethylpyrimdines under alkaline conditions. The mechanism under acidic conditions can be understood in analogy to a Friedel–Crafts alkylation/acylation, which involves a stabilization by the lone electron pair of nitrogen 1 of the positive charge introduced by the alkylating agent.

Under alkaline conditions, an intermittent Michael addition of hydroxide at the 6-position would plausibly generate an enolate-type carbon nucleophile, which

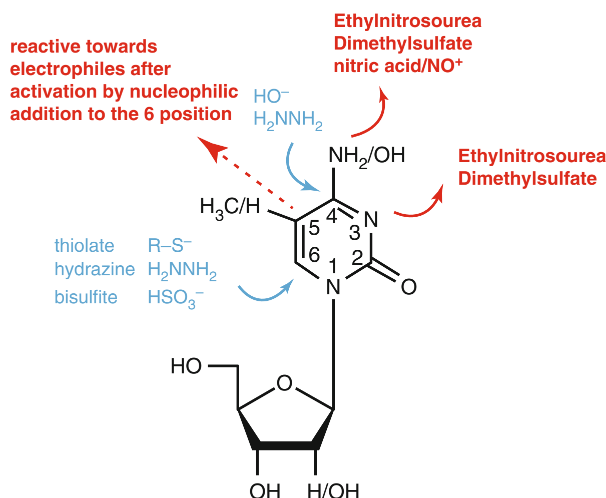


Fig. 2 Chemical reactivity of the pyrimidine C-5 position. Nucleophilic reagents and attack vectors are depicted in *blue*, electrophilic reagents and vectors are in *red*

then reacts with the electrophilic formaldehyde, followed by elimination of the hydroxide to restore the aromatic ring. Note that indeed, the enzymatic mechanisms discussed below for alkylation, acylation, or hydroxymethylation all involve such a Michael attack by a nucleophile, typically a cysteine thiolate (Jurkowski et al. 2008; Motorin et al. 2010). Interestingly, mechanisms discussed for the enzymatic decarboxylation of 5cC and 5fC employ the same path in reverse. Carell et al. described a nonenzymatic in vitro decarboxylation proceeding in the presence of high concentrations of thiol but at low pH (Schiesser et al. 2013; Schiesser et al. 2012). Under the same conditions, removal of formaldehyde from 5hmC was inefficient. Reaction with formaldehyde may be conducted in the presence of amine, resulting in aminomethylation, thus leading to modified pyrimidines that closely resemble native counterparts typically found at 34-position of tRNA. Here again, the catalytic mechanism in the biosynthesis of these modified bases bears similarities (Helm and Alfonzo 2014; El Yacoubi et al. 2012) with that of other modifications using CH_2-THF (*vide infra*).

2 Enzymatic Mechanisms of Pyrimidine Alkylation

Attachment of the alkyl (most frequently -CH_3) group to carbon 5 of pyrimidines U and C can be catalyzed by a variety of enzymes which differ in their origin, sequence, and structure yet employ some common principles of catalysis. At the nucleotide level, this reaction is catalyzed by the extensively studied thymidylate synthase (TS), which is a key target enzyme in certain anticancer and immunosuppressive treatments. TS catalyzes the conversion of dUMP into dTMP, an essential reaction

for the synthesis of DNA nucleotides. At the polynucleotide level, the methylation of C5 in U and C is insured either by specific DNA-MTases (for 5mC formation) of by RNA-specific m^5U-methyltransferases as well as m^5C-methyltransferases.

2.1 The Thymidylate Synthase Family

Thymidylate synthase (TS, EC 2.1.1.45) catalyzes the synthesis of dTMP via a reductive methylation of dUMP. This important enzyme family has been extensively studied for almost 40 years, starting in the late 1970s (Santi 1986; Carreras and Santi 1995). The first characterized enzymes used CH_2-THF as co-substrate, yielding dihydrofolate, from which CH_2-THF was regenerated from serine and $FADH_2$. More recent studies (Koehn and Kohen 2010; Graziani et al. 2006; Agrawal et al. 2004) revealed the existence of a second unusual class of TS, which also act on dUMP and use CH_2-THF, but require $FADH_2$ as a direct reaction cofactor. This family is now called FDTS for flavin-dependent TS. The catalytic mechanism is now established for both enzyme families (Hong et al. 2007; Koehn et al. 2009; Mishanina et al. 2012, 2014).

In the first "classical" family of TS enzymes, the initial step of catalysis relies on a highly conserved Cys residue, which is positioned in the active site of the enzyme. This residue is responsible for the activation of the C5 via addition to the C5 = C6 double bond in the pyrimidine ring, resulting in an enolate intermediate. The enolate's nucleophilic C5 attacks the methylene CH_2 group of the folate co-substrate, forming a covalent ternary complex between the enzyme, dUMP, and the folate. The next step of this reaction is a hydride transfer, which allows the formation of the methylated pyrimidine ring, and is followed by the release of the enzyme via a concerted reaction mechanism corresponding to an elimination that reconstitutes the 5–6 carbon double bond (Islam et al. 2014) (Fig. 3). In the Flavin-dependent TS family, the initial step of the reaction may depend on the enzyme nucleophile (generally an OH-group) or on a direct attachment of $FADH_2$ at the 6-position of the pyrimidine base. In the case of an enzyme nucleophile, the major reaction steps are

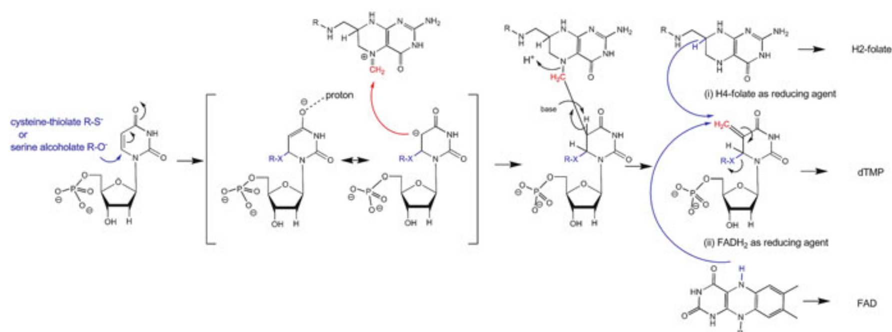


Fig. 3 Enzymatic mechanisms for 2′-desoxythymidine formation in DNA building blocks

rather similar to the "classical" TS, except the last step of hydride transfer, where $FADH_2$ serves as a hydride donor rather than THF. In the case of the direct activation of dUTP by $FADH_2$, after hydride transfer, the $FADH_2$ is replaced by CH_2-THF, and the reaction proceeds by the "classical" way, but without covalent intermediate at the TS active site. The final hydride transfer thus proceeds by an intermolecular rather than an intramolecular reaction (not shown), and THF is the cofactor product as opposed to dihydrofolate in the case of the classical enzymes.

2.2 Enzymes Performing 5-Pyrimidine Methylation in Nucleic Acids

The rT (m^5U) was among the first modified nucleotides discovered in tRNAs, and the respective bacterial enzyme (TrmA or RUMT) catalyzing its biosynthesis was characterized in *E. coli* in the early 1980s (Greenberg and Dudock 1980; Ny and Bjork 1980; Lindstrom et al. 1985). Studies of its enzymatic mechanism identified AdoMet as its CH_3-group donor and Cys324 as a catalytic nucleophile and suggested a simple displacement mechanism of the methylation reaction (Kealey and Santi 1991; Kealey et al. 1991). RUMT catalyzes the modification of U54 in tRNAs, and in addition it is capable of modifying synthetic 16S rRNA *in vitro* (Gu et al. 1994). The yeast homologue of RUMT was also characterized and its tRNA recognition properties studied using synthetic tRNA transcripts (Nordlund et al. 2000; Becker et al. 1997). rT was also found in bacterial rRNA and a different MTase (ygcA, renamed to RumA/RlmD) was found to be responsible for its formation. Mutagenesis of RUMT and structural studies of RumA identified the residues involved in catalysis (Santi and Hardy 1987; Kealey et al. 1994). Bacteria also have an additional enzyme of the same family (RlmC/RumB), catalyzing m^5U747 formation in 23S rRNA. Activity of m^5U-MTases was also detected in Archaea (Constantinesco et al. 1999); however their presence is restricted to the Thermococcales and Nanoarchaeota groups. In *Pyrococcus abyssi*, two close homologues of RlmD fulfill the cellular functions of TrmA (m^5U54 in tRNA) and RlmC (equivalent of m^5U747 in 23S rRNA) (Auxilien et al. 2011). The analysis of m^5U54 formation in *B. subtilis* revealed an unexpected m^5U54-MTase in these gram-positive bacteria. The flavoprotein TrmFO enzyme from *B. subtilis* uses CH_2-THF as a carbon donor, akin to ThyA and ThyX thymidylate synthases (Urbonavicius et al. 2005; Hamdane et al. 2012, 2013). In addition, TrmFO uses the same flavin $FADH_2$ cofactor as the TSFD family, as the reducing agent in the CH_3-group transfer. A similar enzyme was found to catalyze the formation of m^5U1939 in *M. capricolum* 23S rRNA (Lartigue et al. 2014).

The distribution of m^5C in cellular RNAs from different life domains is complex. In bacteria, this modified residue is present in rRNA, but not in other RNA species; in eukaryotes it is found in tRNA, rRNA, and mRNA (Squires et al. 2012; Hussain et al. 2013), while in Archaea its presence seems to be restricted to tRNAs and some sites in mRNAs (Edelheit et al. 2013). Three m^5C residues in *E. coli* rRNA are formed by three specific enzymes, while in yeast three homologues modify both tRNAs and rRNAs. In higher eukaryotes, at least seven or eight specific proteins are

required for the modification of tRNA and cytoplasmic and mitochondrial rRNAs and mRNAs (see Motorin et al. (2010) for further information).

The known enzymes transferring methyl groups from AdoMet to nucleic acids belong to the SPOUT and MTase superfamilies, the latter containing a Rossmann fold for the accommodation of the cofactor. Structure–function relationships in the m^5C-MTase family were combined with bioinformatic analyses (Bujnicki et al. 2004), resulting in a subdivision of the known m^5C-MTases into four major subfamilies: two groups related to Nop2/Nol1 and YebU/Trm4, a large group related to RsmB or Ynl022c, and a small group represented by *P. horikoshi* PH1991 and human NSUN6. Further inspection of homologues in higher eukaryotes (Pavlopoulou and Kossida 2009) suggested the existence of a new subgroup of m^5C-MTase-related proteins, termed RCMT9, with members distantly related to Trm4 and a distribution restricted to four taxons. A detailed discussion of the distribution of m^5C-forming enzymes in the different kingdoms is given in Motorin et al. (2010).

Higher eukaryotes also have another distinct family of m^5C-RNA-MTAses derived from former m^5C:DNA-MTases (DNMT2-related family). These enzymes have different catalytic mechanisms but evolved to modify tRNAs at 38-position (Goll et al. 2006). For information on m^5C-MTases acting on DNA, i.e., enzymes of the DNMT family, see elsewhere in this book.

2.3 Catalytic Mechanisms in the Formation of rT, m^5C, and 5mC

The catalytic strategies employed for the alkylation of the carbon 5 in pyrimidines share some common elements, which derive from the heterocycle reactivity, as outlined above. Some basic elements already appeared in the discussion of the thymidylate synthase in above. In all cases, the Michael addition of an anionic nucleophile to the 6-position of the pyrimidine ring produces a nucleophilic carbon with partial carbanion character at 5-position (Fig. 4). In uridines, the Michael addition produces an intermediate, in which the negative charge is delocalized in an enolate structure. Arguably, this intermediate might be stabilized by a hydrogen bond of the enolate oxygen before reacting as a nucleophile with the carbon electrophile provided as a cofactor in the form of AdoMet or CH_2-THF. In cytidine substrates, the mechanisms comprise an enamine intermediate instead of an enolate, and the mechanisms discussed in literature typically include an acidic residue in the catalytic site, which may intermittently protonate nitrogen 3 to stabilize this enamine intermediate. So far, the known enzymes acting on cytidines in both RNA and DNA exclusively use AdoMet as an electrophilic carbon source, while rT can be formed from either AdoMet or CH_2-THF. This has interesting implications, namely, (i) that a cytidine methyltransferase using CH_2-THF might so far have eluded detection and (ii) that the formation of rT has been invented multiple times with different cofactors in the course of evolution, in particular at 54-position of tRNA (Hamdane et al. 2012; Nordlund et al. 2000; Ny and Bjork 1980). Furthermore, with an eye to the more sophisticated U34 modifications occurring in tRNA, we note that a covalent

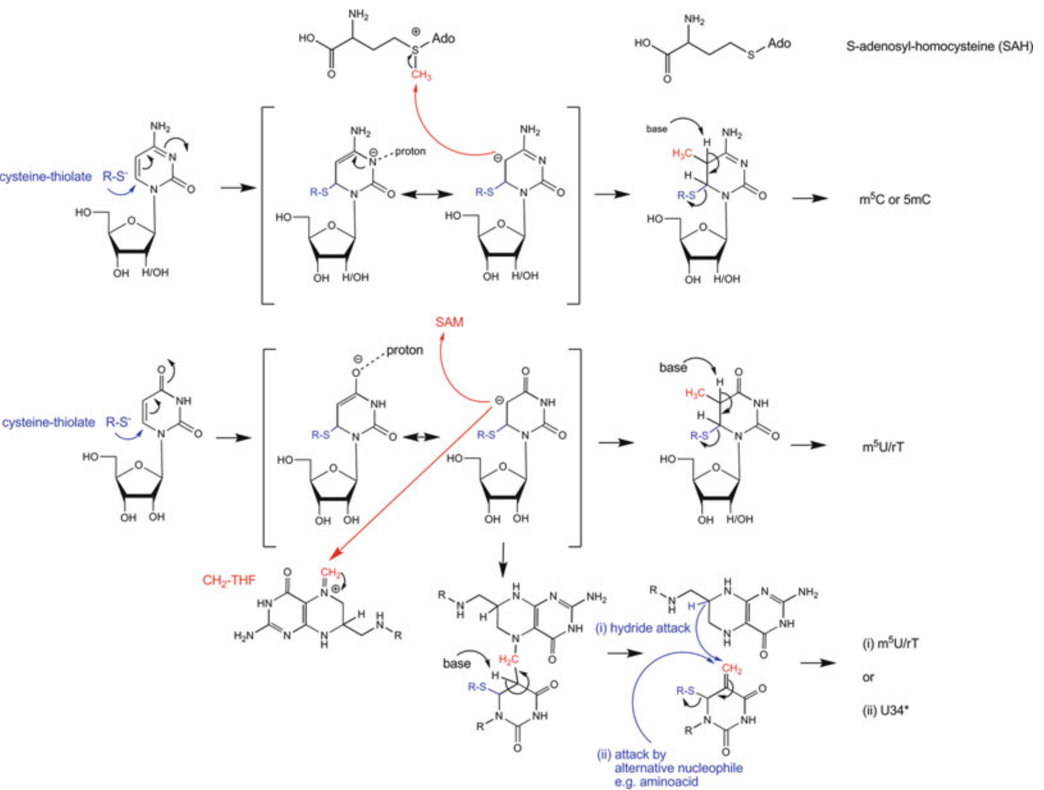


Fig. 4 Posttranscriptional ribouridine and cytidine methylation with the AdoMet cofactor. The first step uses a catalytic cysteine in the enzyme active site for activation of the C5=C6 bond via formation of a covalent enzyme-RNA enolate intermediate. This allows the transfer of the CH_3-group from AdoMet to the 5-position of the base, followed by deprotonation via Glu358 and the enzyme release from the covalent enzyme–RNA complex (Reviewed in Kealey et al. 1994)

enzyme-thiol-pyrimidine-methylene-folate intermediate (as reported/postulated for TS (Fig. 3) and TrmFO) can not only be resolved by a reduction with a hydride equivalent from the folate (Fig. 4(i)) but also with other nucleophiles (Helm and Alfonzo 2014) (Fig. 4 (ii)). The aromaticity of the pyrimidine base is restored in an elimination step featuring an abstraction of a proton from the C5, which regenerates the thiolate used in the initial activation step via Michael addition. In a number of RNA m^5C:MTases, a second cysteine was reported to be crucial to this regeneration (reviewed in Motorin et al. (2010)).

In the framework of the above-described common elements, the various enzymes differ from one another by the amino acids that embody the different roles outlined above, such as activating nucleophile, general acid, general base, etc. In addition, the position of such residues, while relatively conserved in the spatial arrangement of the active site, may vary within the polypeptide sequence of the different enzymes. This has been especially well studied in m^5C:MTases, where these residues are located within several conserved motives numbered I through X, which, in their order of appearance in the primary sequence, undergo permutation among the different enzymes of the bacterial and plant DNA-MTase family and some of the RNA-m^5C:MTase (reviewed in Motorin et al. (2010)).

2.4 Catalytic Mechanisms in the Formation of Exotic U34 Modifications

As already mentioned several times, the most bewildering variety of 5-pyrimdine modifications are found at uridine 34 in the anticodon of tRNAs, where they play a crucial role in mRNA decoding. Alkylations predominate among these modifications, and some of the enzymatic mechanisms bear strong similarity with those applied in simple methylations. For example, certain enzymes use CH_2-THF to transfer a formaldehyde equivalent to the C5, and instead of reducing it to the methyl group using a hydride donor, such as tetrahydrofolate (H4-folate) or $FADH_2$ (Fig. 4), the amino group of certain amino acids such as taurine or glycine serves as the attacking nucleophile, leading to the structures displayed in Fig. 1c, which correspond to the overall product of a Mannich reaction (Helm and Alfonzo 2014). Recently, two groups discovered new types of reactive intermediates formed and employed in catalytic mechanisms of U34 modification. The Almo group reported the conversion of the conventional AdoMet into a novel derivative, carboxy-*S*-adenosyl-L-methionine, which is used by the bacterial CmoB enzyme to introduce the carboxymethyl into 5-hydroxyuridine (ho5U), yielding a 5-oxyacetyluridine (cmo^5U) (Kim et al. 2013, 2015). Most interestingly, Huang's group reported the use of AdoMet by a radical AdoMet enzyme from the elongator complex to generate a radical from the methyl group of the acetic acid moiety in acetyl CoA, which would then add to the C5–C6 double bond of uridine34 in Archaea and Eukarya (Selvadurai et al. 2014).

3 Functions of Alkylated Pyrimidine Nucleosides

In view of the plethora of different structures of alkylated pyrimidines already discovered, it is clear that there cannot be one function common to all of them. Indeed, new facets of functions are being continuously discovered in very diverse areas of molecular life sciences, and since this is not the focus of this chapter, we will only provide references to few known functions. One common biophysical property of 5-methylpyrimdines is that they enhance stacking in A- and B-helices of nucleic acids, leading to a structural stabilization that is typically reflected in an increased thermal stability detected in melting experiments. This applies to thymidine and 5-methylcytidine in DNA and RNA alike. The role of 5-methylcytidine in mammalian epigenetics, as well as in the restriction/methylation systems of bacteria, is detailed elsewhere in this book. Of interest, certain bacteriophages use particular 5-pyrimdmine modifications to escape bacterial restriction (Gommers-Ampt and Borst 1995). Curiously, the roles of ribothymidine and 5-methylribocytidine in RNA have remained little understood despite their long-standing tenure in the zoo of known RNA modifications (Motorin and Helm 2011). This is likely due to the fact that their principle occurrences in tRNA and rRNA concern heavily modified RNAs, where a plethora of modifications cooperate in a network fashion to modulate RNA activity (Motorin and Helm 2010). The role of U34 modifications in tRNA has already been alluded to, although the generic explanation of mRNA decoding on the ribosome does not do justice to the plethora of structures found here. Apparently, there is no universally perfect modification at this site that suits all organisms, and the variety of conditions under which protein synthesis must take place has led to the emergence of numerous chemical solutions in different species. Along this line, the recent findings that tRNA anticodon modifications are dynamically responding to stress conditions point to an especially sensitive environment that is subject to constant tuning and further evolution.

References

Agrawal N, Lesley SA, Kuhn P, Kohen A. Mechanistic studies of a flavin-dependent thymidylate synthase. Biochemistry. 2004;43(32):10295–301. doi:10.1021/bi0490439.

Auxilien S, Rasmussen A, Rose S, Brochier-Armanet C, Husson C, Fourmy D, et al. Specificity shifts in the rRNA and tRNA nucleotide targets of archaeal and bacterial m5U methyltransferases. RNA. 2011;17(1):45–53. doi:10.1261/rna.2323411.

Becker HF, Motorin Y, Sissler M, Florentz C, Grosjean H. Major identity determinants for enzymatic formation of ribothymidine and pseudouridine in the T psi-loop of yeast tRNAs. J Mol Biol. 1997;274(4):505–18. doi:10.1006/jmbi.1997.1417.

Bogdanovic O, Gomez-Skarmeta JL. Embryonic DNA methylation: insights from the genomics era. Brief Funct Genomics. 2014;13(2):121–30. doi:10.1093/bfgp/elt039.

Bujnicki JM, Feder M, Ayres CL, Redman KL. Sequence-structure-function studies of tRNA:m5C methyltransferase Trm4p and its relationship to DNA:m5C and RNA:m5U methyltransferases. Nucleic Acids Res. 2004;32(8):2453–63. doi:10.1093/nar/gkh564.

Burgess AL, David R, Searle IR. Conservation of tRNA and rRNA 5-methylcytosine in the kingdom Plantae. BMC Plant Biol. 2015;15:199. doi:10.1186/s12870-015-0580-8.

Carreras CW, Santi DV. The catalytic mechanism and structure of thymidylate synthase. Annu Rev Biochem. 1995;64:721–62. doi:10.1146/annurev.bi.64.070195.003445.

Constantinesco F, Motorin Y, Grosjean H. Transfer RNA modification enzymes from Pyrococcus furiosus: detection of the enzymatic activities in vitro. Nucleic Acids Res. 1999;27(5):1308–15.

Conte MR, Galeone A, Avizonis D, Hsu VL, Mayol L, Kearns DR. Solid phase synthesis of 5-hydroxymethyluracil containing DNA. Bioorg Med Chem Lett. 1992;2:79–81.

Edelheit S, Schwartz S, Mumbach MR, Wurtzel O, Sorek R. Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. PLoS Genet. 2013;9(6):e1003602. doi:10.1371/journal.pgen.1003602.

El Yacoubi B, Bailly M, de Crecy-Lagard V. Biosynthesis and function of posttranscriptional modifications of transfer RNAs. Annu Rev Genet. 2012;46:69–95. doi:10.1146/annurev-genet-110711-155641.

Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci U S A. 1992;89(5):1827–31.

Fu L, Guerrero CR, Zhong N, Amato NJ, Liu Y, Liu S, et al. Tet-mediated formation of 5-hydroxymethylcytosine in RNA. J Am Chem Soc. 2014;136(33):11582–5. doi:10.1021/ja505305z.

Giege R, Helm M, Florentz C. Chemical and enzymatic Probing of RNA Structure. Comprehensive Natural Product Chemistry. In: Söll D, Nishimura S (editors). Oxford: Pergamon Press; 1999;6:63–77.

Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh CL, Zhang X, et al. Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science. 2006;311(5759):395–8. doi:10.1126/science.1120976.

Gommers-Ampt JH, Borst P. Hypermodified bases in DNA. FASEB J Off Publ Fed Am Soc Exp Biol. 1995;9(11):1034–42.

Graziani S, Bernauer J, Skouloubris S, Graille M, Zhou CZ, Marchand C, et al. Catalytic mechanism and structure of viral flavin-dependent thymidylate synthase ThyX. J Biol Chem. 2006;281(33):24048–57. doi:10.1074/jbc.M600745200.

Greenberg R, Dudock B. Isolation and characterization of m5U-methyltransferase from Escherichia coli. J Biol Chem. 1980;255(17):8296–302.

Gu X, Ofengand J, Santi DV. In vitro methylation of Escherichia coli 16S rRNA by tRNA (m5U54)-methyltransferase. Biochemistry. 1994;33(8):2255–61.

Hamdane D, Argentini M, Cornu D, Golinelli-Pimpaneau B, Fontecave M. FAD/folate-dependent tRNA methyltransferase: flavin as a new methyl-transfer agent. J Am Chem Soc. 2012;134(48):19739–45. doi:10.1021/ja308145p.

Hamdane D, Bruch E, Un S, Field M, Fontecave M. Activation of a unique flavin-dependent tRNA-methylating agent. Biochemistry. 2013;52(49):8949–56. doi:10.1021/bi4013879.

He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science. 2011;333(6047):1303–7. doi:10.1126/science.1210944.

Helm M, Alfonzo JD. Posttranscriptional RNA Modifications: playing metabolic games in a cell's chemical Legoland. Chem Biol. 2014;21(2):174–85. doi:10.1016/j.chembiol.2013.10.015.

Hong B, Maley F, Kohen A. Role of Y94 in proton and hydride transfers catalyzed by thymidylate synthase. Biochemistry. 2007;46(49):14188–97. doi:10.1021/bi701363s.

Hussain S, Aleksic J, Blanco S, Dietmann S, Frye M. Characterizing 5-methylcytosine in the mammalian epitranscriptome. Genome Biol. 2013;14(11):215. doi:10.1186/gb4143.

Islam Z, Strutzenberg TS, Gurevic I, Kohen A. Concerted versus stepwise mechanism in thymidylate synthase. J Am Chem Soc. 2014;136(28):9850–3. doi:10.1021/ja504341g.

Jurkowski TP, Meusburger M, Phalke S, Helm M, Nellen W, Reuter G, et al. Human DNMT2 methylates tRNA(Asp) molecules using a DNA methyltransferase-like catalytic mechanism. RNA. 2008;14(8):1663–70. doi:10.1261/rna.970408.

Kealey JT, Santi DV. Identification of the catalytic nucleophile of tRNA (m5U54)methyltransferase. Biochemistry. 1991;30(40):9724–8.
Kealey JT, Lee S, Floss HG, Santi DV. Stereochemistry of methyl transfer catalyzed by tRNA (m5U54)-methyltransferase--evidence for a single displacement mechanism. Nucleic Acids Res. 1991;19(23):6465–8.
Kealey JT, Gu X, Santi DV. Enzymatic mechanism of tRNA (m5U54)methyltransferase. Biochimie. 1994;76(12):1133–42.
Khursid M, Khan A, Qudrat-E-Khuda M. Synthesis of 5-Hydroxymethylcytidine. J Chem Soc Pak. 1982;4(3):167–8.
Kim J, Xiao H, Bonanno JB, Kalyanaraman C, Brown S, Tang X, et al. Structure-guided discovery of the metabolite carboxy-SAM that modulates tRNA function. Nature. 2013;498(7452):123–6. doi:10.1038/nature12180.
Kim J, Xiao H, Koh J, Wang Y, Bonanno JB, Thomas K, et al. Determinants of the CmoB carboxymethyl transferase utilized for selective tRNA wobble modification. Nucleic Acids Res. 2015;43(9):4602–13. doi:10.1093/nar/gkv206.
Koehn EM, Kohen A. Flavin-dependent thymidylate synthase: a novel pathway towards thymine. Arch Biochem Biophys. 2010;493(1):96–102. doi:10.1016/j.abb.2009.07.016.
Koehn EM, Fleischmann T, Conrad JA, Palfey BA, Lesley SA, Mathews II, et al. An unusual mechanism of thymidylate biosynthesis in organisms containing the thyX gene. Nature. 2009;458(7240):919–23. doi:10.1038/nature07973.
Kong LQ, Zhao J, Fan L, Yang DC. The facile synthetic method of 5-hydroxymethyluracil. Chin Chem Lett. 2009;20(3):314–6.
Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science. 2009;324(5929):929–30. doi:10.1126/science.1169786.
Kun A, Szilagyi A, Konnyu B, Boza G, Zachar I, Szathmary E. The dynamics of the RNA world: insights and challenges. Ann N Y Acad Sci. 2015;1341:75–95. doi:10.1111/nyas.12700.
Lartigue C, Lebaudy A, Blanchard A, El Yacoubi B, Rose S, Grosjean H, et al. The flavoprotein Mcap0476 (RlmFO) catalyzes m5U1939 modification in Mycoplasma capricolum 23S rRNA. Nucleic Acids Res. 2014;42(12):8073–82. doi:10.1093/nar/gku518.
Lindstrom PH, Stuber D, Bjork GR. Genetic organization and transcription from the gene (trmA) responsible for synthesis of tRNA (uracil-5)-methyltransferase by Escherichia coli. J Bacteriol. 1985;164(3):1117–23.
Machnicka MA, Milanowska K, Osman Oglou O, Purta E, Kurkowska M, Olchowik A, et al. MODOMICS: a database of RNA modification pathways--2013 update. Nucleic Acids Res. 2013;41(Database issue):D262–7. doi:10.1093/nar/gks1007.
Militello KT, Chen LM, Ackerman SE, Mandarano AH, Valentine EL. A map of 5-methylcytosine residues in Trypanosoma brucei tRNA revealed by sodium bisulfite sequencing. Mol Biochem Parasitol. 2014;193(2):122–6. doi:10.1016/j.molbiopara.2013.12.003.
Mishanina TV, Koehn EM, Conrad JA, Palfey BA, Lesley SA, Kohen A. Trapping of an intermediate in the reaction catalyzed by flavin-dependent thymidylate synthase. J Am Chem Soc. 2012;134(9):4442–8. doi:10.1021/ja2120822.
Mishanina TV, Corcoran JM, Kohen A. Substrate activation in flavin-dependent thymidylate synthase. J Am Chem Soc. 2014;136(30):10597–600. doi:10.1021/ja506108b.
Moriya J, Yokogawa T, Wakita K, Ueda T, Nishikawa K, Crain PF, et al. A novel modified nucleoside found at the first position of the anticodon of methionine tRNA from bovine liver mitochondria. Biochemistry. 1994;33(8):2234–9.
Motorin Y, Helm M. tRNA stabilization by modified nucleotides. Biochemistry. 2010;49(24):4934–44. doi:10.1021/bi100408z.
Motorin Y, Helm M. RNA nucleotide methylation. Wiley Interdiscip Rev RNA. 2011;2(5):611–31. doi:10.1002/wrna.79.
Motorin Y, Lyko F, Helm M. 5-methylcytosine in RNA: detection, enzymatic formation and biological functions. Nucleic Acids Res. 2010;38(5):1415–30. doi:10.1093/nar/gkp1117.
Muller UF. Re-creating an RNA world. Cell Mol Life Sci. 2006;63(11):1278–93. doi:10.1007/s00018-006-6047-1.

Nordlund ME, Johansson JO, von Pawel-Rammingen U, Bystrom AS. Identification of the TRM2 gene encoding the tRNA(m5U54)methyltransferase of Saccharomyces cerevisiae. RNA. 2000;6(6):844–60.

Ny T, Bjork GR. Cloning and restriction mapping of the trmA gene coding for transfer ribonucleic acid (5-methyluridine)-methyltransferase in Escherichia coli K-12. J Bacteriol. 1980;142(2):371–9.

Pavlopoulou A, Kossida S. Phylogenetic analysis of the eukaryotic RNA (cytosine-5)-methyltransferases. Genomics. 2009;93(4):350–7. doi:10.1016/j.ygeno.2008.12.004.

Pfaffeneder T, Hackner B, Truss M, Munzel M, Muller M, Deiml CA, et al. The discovery of 5-formylcytosine in embryonic stem cell DNA. Angewandte Chemie. 2011;50(31):7008–12. doi:10.1002/anie.201103899.

Pfaffeneder T, Spada F, Wagner M, Brandmayr C, Laube SK, Eisen D, et al. Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. Nat Chem Biol. 2014;10(7):574–81. doi:10.1038/nchembio.1532.

Racz I, Kiraly I, Lasztily D. Effect of light on the nucleotide composition of rRNA of wheat seedlings. Planta. 1978;142(3):263–7. doi:10.1007/BF00385075.

Roberts RJ, Vincze T, Posfai J, Macelis D. REBASE--restriction enzymes and DNA methyltransferases. Nucleic Acids Res. 2005;33(Database issue):D230–2. doi:10.1093/nar/gki029.

Santi DV. The mechanism and structure of thymidylate synthetase. Nucleic Acids Symp Ser. 1986;17:125–6.

Santi DV, Hardy LW. Catalytic mechanism and inhibition of tRNA (uracil-5-)methyltransferase: evidence for covalent catalysis. Biochemistry. 1987;26(26):8599–606.

Schaefer M, Pollex T, Hanna K, Lyko F. RNA cytosine methylation analysis by bisulfite sequencing. Nucleic Acids Res. 2009;37(2):e12. doi:10.1093/nar/gkn954.

Schiesser S, Hackner B, Pfaffeneder T, Muller M, Hagemeier C, Truss M, et al. Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. Angewandte Chemie. 2012;51(26):6516–20. doi:10.1002/anie.201202583.

Schiesser S, Pfaffeneder T, Sadeghian K, Hackner B, Steigenberger B, Schroder AS, et al. Deamination, oxidation, and C-C bond cleavage reactivity of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine. J Am Chem Soc. 2013;135(39):14593–9. doi:10.1021/ja403229y.

Selvadurai K, Wang P, Seimetz J, Huang RH. Archaeal Elp3 catalyzes tRNA wobble uridine modification at C5 via a radical mechanism. Nat Chem Biol. 2014;10(10):810–2. doi:10.1038/nchembio.1610.

Squires JE, Patel HR, Nousch M, Sibbritt T, Humphreys DT, Parker BJ, et al. Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. Nucleic Acids Res. 2012;40(11):5023–33. doi:10.1093/nar/gks144.

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2009;324(5929):930–5. doi:10.1126/science.1170116.

Urbonavicius J, Skouloubris S, Myllykallio H, Grosjean H. Identification of a novel gene encoding a flavin-dependent tRNA:m5U methyltransferase in bacteria--evolutionary implications. Nucleic Acids Res. 2005;33(13):3955–64. doi:10.1093/nar/gki703.

Vaught JD, Dewey T, Eaton BE. T7 RNA polymerase transcription with 5-position modified UTP derivatives. J Am Chem Soc. 2004;126(36):11231–7. doi:10.1021/ja049009h.

Vaught JD, Bock C, Carter J, Fitzwater T, Otis M, Schneider D, et al. Expanding the chemistry of DNA for in vitro selection. J Am Chem Soc. 2010;132(12):4141–51. doi:10.1021/ja908035g.

Wyatt GR, Cohen SS. The bases of the nucleic acids of some bacterial and animal viruses: the occurrence of 5-hydroxymethylcytosine. Biochem J. 1953;55(5):774–82.

Bacterial DNA Methylation and Methylomes

Josep Casadesús

Abstract

Formation of C5-methylcytosine, N4-methylcytosine, and N6-methyladenine in bacterial genomes is postreplicative and involves transfer of a methyl group from S-adenosyl-methionine to a base embedded in a specific DNA sequence context. Most bacterial DNA methyltransferases belong to restriction-modification systems; in addition, "solitary" or "orphan" DNA methyltransferases are frequently found in the genomes of bacteria and phage. Base methylation can affect the interaction of DNA-binding proteins with their cognate sites, either by a direct effect (e.g., steric hindrance) or by changes in DNA topology. In both *Alphaproteobacteria* and *Gammaproteobacteria*, the roles of DNA base methylation are especially well known for N6-methyladenine, including control of chromosome replication, nucleoid segregation, postreplicative correction of DNA mismatches, cell cycle-coupled transcription, formation of bacterial cell lineages, and regulation of bacterial virulence. Technical procedures that permit genome-wide analysis of DNA methylation are nowadays expanding our knowledge of the extent, evolution, and physiological significance of bacterial DNA methylation.

Abbreviations

AdoMet	S-Adenosyl-L-methionine
CcrM	Cell cycle-regulated methylase
Cori	Replication origin of the *Caulobacter* chromosome
CRISP-R	Clustered regularly interspaced short palindromic repeats

J. Casadesús
Departamento de Genética, Universidad de Sevilla, Apartado 1095, Seville 41080, Spain
e-mail: casadesus@us.es

© Springer International Publishing Switzerland 2016

A. Jeltsch, R.Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*,
Advances in Experimental Medicine and Biology 945,
DOI 10.1007/978-3-319-43624-1_3

IPD	Interpulse duration
LD_{50}	Median lethal dose
Mod	Modification gene in restriction-modification systems
oriC	Replication origin of the *E. coli* chromosome
SMALR	Single-molecule modification analysis of long reads
SMRT	Single-molecule real time
SPI-1	*Salmonella* pathogenicity island 1
UAS	Upstream regulatory region
VSP	Very-short-patch

1 Introduction

Bacterial genomes contain small amounts of N4-methylcytosine, C5-methylcytosine, and N6-methyladenine (Vanyushin et al. 1968; Cheng 1995; Jeltsch 2002). Base methylation is catalyzed by DNA methyltransferases that recognize specific DNA motifs, and it occurs after DNA replication. The majority of DNA methyltransferases described in the literature are part of restriction-modification systems, each made of a restriction endonuclease and an adenine or cytosine DNA methyltransferase. In the 1960s, restriction-modification of DNA provided an explanation for an enigmatic phenomenon described in the previous decade: the modification of bacteriophage host range upon passage through specific host strains (Bertani and Weigle 1953). Growth of virulent phages on bacterial cells was found to be restricted by endonucleases that attack nonmethylated phage DNA, while host DNA is protected by a specific methylation pattern at adenine or cytosine moieties (Arber and Linn 1969).

Restriction-modification systems are classified into three types on the basis of structural features, pattern of DNA cleavage, and cofactor requirements (Wilson and Murray 1991; Loenen et al. 2014) (Table 1). In types I and III, the DNA adenine or

Table 1 Classification of restriction-modification systems

Type	Composition	Cofactors	DNA restriction pattern
I	Multiple subunits	Mg^{++} AdoMet ATP	Random cleavage far from asymmetrical recognition sites
II	Separate enzymes for restriction and modification	Mg^{++}	Cleavage within symmetrical recognition sequences
IIs	Separate enzymes for restriction and modification	Mg^{++}	Cleavage at fixed distance from symmetrical recognition sequences
III	Separate enzymes for restriction and modification	Mg^{++} ATP	Cleavage at fixed distance from symmetrical recognition sequences
IV	Several subunits	Mg^{++} GTP	Cleavage of DNA containing methylated nucleotides

Additional information in the reviews by Wilson and Murray (1991), Jeltsch (2002), and Loenen et al. (2014)

cytosine methyltransferase is part of a multisubunit protein complex involved in both restriction and modification. In contrast, type II systems consist of two separate enzymes, a restriction endonuclease and an adenine or cytosine DNA methyltransferase. In addition to these three types of restriction-modification systems in which DNA methylation protects against endonucleolytic cleavage, restriction systems specific for methylated bases (type IV) have also been described (Wilson and Murray 1991; Loenen and Raleigh 2014). In the last decade, a bacterial immunity system based on clustered regularly interspaced short palindromic repeats (CRISP-R) has revealed an additional unsuspected mechanism of defense against virulent phages, but this process does not involve DNA methylation (Bhaya et al. 2012).

Aside from providing barriers against foreign DNA invasion, restriction-modification systems may play additional roles in the bacterial world (Vasu and Nagaraja 2013). For instance, incomplete protection of host DNA in a bacterial population may permit rare events of acquisition of foreign DNA by bacterial cells in a sort of evolutionary bet hedging (Arber 2000). Furthermore, restriction-modification systems may contribute to maintain the identity of bacterial lineages: in *Neisseria meningitidis*, commensal isolates harbor a DNA adenine methyltransferase that methylates 5′-GATC-3′ sites, while pathogenic isolates produce a restriction endonuclease that cleaves 5′-GATC-3′ sites. DNA transfer from commensal to pathogenic isolates is thus prevented (Jeltsch 2003).

In addition to restriction-modification systems, bacterial and phage genomes encode DNA methyltransferases that do not have a restriction enzyme counterpart and are known as "solitary" or "orphan" DNA methyltransferases (Vanyushin et al. 1971; Marinus 1996; Løbner-Olesen et al. 2005; Wion and Casadesus 2006). Solitary methyltransferases may have derived from ancestral restriction-modification systems that lost their restriction enzyme component. In support of this view, restriction-modification systems in which the modification enzyme is functional but the restriction enzyme is inactive exist in *Helicobacter pylori* and probably in other bacterial species (Fox et al. 2007a). However, the abundance of solitary DNA methyltransferases in bacterial and archaebacterial genomes raises the alternative possibility that DNA methylation may be an ancestral trait maintained by natural selection (Blow et al. 2016). If this view is correct, restriction enzymes might have evolved to ensure that the methyltransferases remained active rather than to confer protection against bacteriophages and other infectious DNA molecules (Blow et al. 2016). In fact, an analogy between restriction-modification systems and addiction modules has been drawn (Naito et al. 1995). The selective value of DNA methyltransferases may also be supported by their occurrence in bacterial species with small genomes (Lluch-Senar et al. 2013).

The roles played by DNA methylation in bacterial physiology have been mostly investigated in two model methyltransferases, the Dam methylase of *Gammaproteobacteria* and the CcrM methylase of *Alphaproteobacteria* (Wion and Casadesus 2006; Collier 2009; Sanchez-Romero et al. 2015; Mohapatra et al. 2014). The marks introduced into DNA by these methylases provide signals for a variety of physiological processes including regulation of the cell cycle and epigenetic control of gene expression. Traditionally, epigenetic regulation was considered an exclusive

task of solitary DNA methyltransferases like Dam and CcrM and a consequence of long coevolution that adapted the epigenome to physiological needs (Casadesus and Low 2006). However, this view has been challenged by the finding that a prophage-encoded DNA methyltransferase belonging to a restriction-modification system controls gene expression (Fang et al. 2012). The need of long coevolution to integrate DNA methylation into host regulatory circuits is likewise challenged by a study showing that a DNA methyltransferase acquired by horizontal transfer can overtake the control of housekeeping functions in *Vibrio cholerae* (Chao et al. 2015). These recent and exciting findings, together with the development of new analytical technologies, may bring about novel paradigms on bacterial DNA methylation. In the meantime, the DNA adenine methyltransferases Dam and CcrM remain classical examples, and for this reason they receive close attention in this chapter.

2 Solitary DNA Methyltransferases in Bacteria

2.1 Dam Methyltransferase

The DNA adenine methyltransferase known as Dam methylase, which is present in multiple genera of *Gammaproteobacteria*, was initially characterized in *E. coli* (Marinus 1996). Dam is a monomer in solution and catalyzes the transfer of a methyl group from *S*-adenosyl-L-methionine (AdoMet) to the N6 position of the adenine residue in 5′-GATC-3′ sequences (Marinus 1996). Although the natural substrate for the enzyme is hemimethylated DNA formed upon DNA replication, nonmethylated GATC sites are also Dam substrates. Hence, Dam is both a de novo methylase and a maintenance methylase. In fact, there is little difference in the rate of methylation between nonmethylated and hemimethylated DNA (Herman and Modrich 1982). Dam contains two AdoMet binding sites: one is the catalytic site and the other increases specific binding to DNA, probably through an allosteric transformation (Bergerat et al. 1991). Dam is a highly processive enzyme and methylates about 55 GATC sites per binding event (Urig et al. 2002).

Expression of the Dam methylase is under transcriptional control in *E. coli*. The *dam* gene is driven by five promoters (Løbner-Olesen et al. 1992), and one of them is regulated by the growth rate. This control adjusts the cellular level of Dam, which consists of about 130 molecules during exponential growth in rich medium. The level of Dam methylase may be additionally controlled by proteolysis (Calmann and Marinus 2003).

2.2 CcrM Methylase

CcrM (acronym for "cell cycle-regulated methylase") was initially identified in *Caulobacter crescentus* (Stephens et al. 1996; Robertson et al. 2000; Kahng and Shapiro 2001; Mohapatra et al. 2014). The target for CcrM is 5′-GANTC-3′, where "N" is any nucleotide. CcrM is active both as a monomer and as a dimer, uses

AdoMet as methyl donor, and shows a slight preference for hemimethylated DNA substrates (Stephens et al. 1996; Robertson et al. 2000; Kahng and Shapiro 2001; Mohapatra et al. 2014). A mechanistic difference between CcrM and Dam is that CcrM is not a processive enzyme (Albu et al. 2011). Additional differences are that Dam is present in the cell throughout the cell cycle, while the expression of CcrM is restricted to a late stage of chromosome replication (Stephens et al. 1996) and that CcrM is essential in rich medium (Gonzalez and Collier 2013), while Dam is not (Marinus 1996).

CcrM homologs have been found in *Agrobacterium tumefaciens*, the causative agent of crown gall disease in dicotyledonous plants (Kahng and Shapiro 2001); *Rhizobium meliloti*, the nitrogen-fixing symbiont of legumes (Wright et al. 1997); and in the animal pathogen *Brucella abortus* (Robertson et al. 2000). In *Brucella*, aberrant CcrM expression impairs proliferation in murine macrophages, suggesting a role in pathogenesis (Robertson et al. 2000).

A gene that encodes a CcrM homolog known as YhdJ is found in the genomes of *E. coli* and *Salmonella* (Broadbent et al. 2007). YhdJ can methylate the 3′ adenosine moiety of 5′-AGTCAT-3′ targets in vitro. However, YhdJ does not seem to be expressed in vivo, at least under laboratory conditions (Broadbent et al. 2007).

2.3 Dcm

The Dcm methyltransferase was described in *E. coli* several decades ago and is also present in other enteric bacteria (Marinus 1996). Dcm methylates the C5 position of internal cytosine residues in 5′-CCAGG-3′ and 5′-CCTGG-3′ sites. Bioinformatic analysis suggests that Dcm may be a protein of ~53 kD (Marinus 1996).

3 Role of DNA Methylation in DNA Mismatch Repair

DNA adenine methylase mutants of *E. coli* and *S. enterica* show a hypermutable phenotype with excess transition mutations (purine to purine or pyrimidine to pyrimidine) (Marinus and Morris 1974; Torreblanca and Casadesus 1996; Glickman et al. 1978). This phenotype is indicative of the inability of *dam* mutants to repair DNA replication errors that introduce mismatched base pairs. Because replication-generated DNA mismatches involve normal (non-damaged) nucleotides, they cannot be repaired by base excision repair or by nucleotide excision repair. Furthermore, the repair machinery needs to discriminate between the error-free template strand and the error-prone daughter strand. *E. coli* and other *Gammaproteobacteria* use Dam hemimethylation for this discrimination (Fig. 1). Mismatched base pairs are recognized by a protein known as MutS, which then recruits two additional proteins, MutL and MutH (Iyer et al. 2006). When the MutS-MutL-MutH ternary complex is assembled at a DNA mismatch, MutH acquires endonuclease activity and cleaves the phosphodiester bond of the nonmethylated DNA strand located at the 5′ side of the G in the closest GATC. After cleavage, the UvrD helicase dislodges

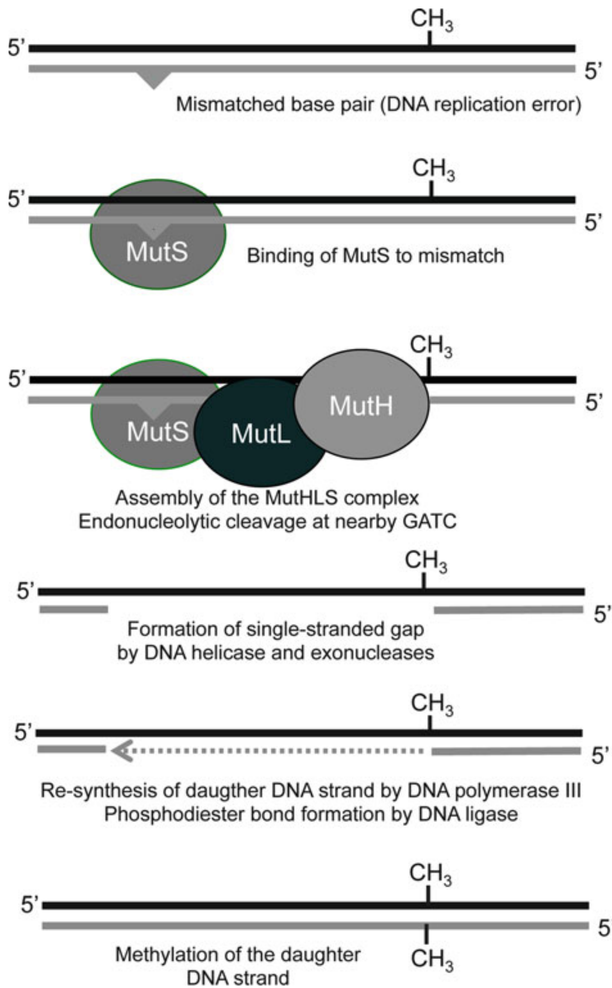


Fig. 1 Dam-directed repair of a DNA mismatch. Detection of the mismatch by MutS recruits MutH and MutL, and the MutHLS complex is assembled. MutH-mediated endonucleolytic cleavage of the newly synthesized DNA strand occurs at the nearest GATC site. Depending on the distance from the mismatch to the GATC, cleavage may require DNA looping (not drawn)

MutH from the ternary complex, and DNA unwinding is followed by single-strand degradation by exonucleases. The resulting gap is then filled in by DNA polymerase III, and the nick is sealed by formation of a phosphodiester bond by DNA ligase. Finally, Dam methylase converts the hemimethylated GATC to a fully methylated site (Fig. 1). Because MutH cannot cleave methylated DNA, mismatch repair is confined to a short hemimethylated DNA region, probably around 10 kb long, just behind the replication fork. Transient lack of GATC methylation in the newly synthesized strand thus provides the signal for DNA strand discrimination by MutHLS (Pukkila et al. 1983).

Upon assembly of the MutH-MutL-MutS complex, MutH endonuclease can cleave nonmethylated GATC duplexes. As a consequence, *dam* mutants can suffer MutH-mediated cleavage in both DNA strands (Marinus and Casadesus 2009). Double-strand DNA breakage explains several traits of *dam* mutants: (i) sensitivity to agents that induce DNA injuries recognized by MutS, (ii) dependence on homologous recombination and other DNA repair functions to cope with DNA damage, and (iii) permanent induction of the SOS response (Marinus 1996; Marinus and Casadesus 2009).

E. coli and *Salmonella* strains that overproduce Dam methylase show even higher mutation rates than *dam* mutants (Pukkila et al. 1983; Marinus et al. 1984; Torreblanca and Casadesus 1996). Albeit paradoxical at first sight, this observation underscores the relevance of transient hemimethylation as a signal for mismatch repair: excess Dam methylase shortens the hemimethylation period in newly replicated DNA molecules, thus preventing MutH-mediated GATC cleavage of the daughter DNA strand in the vicinity of mismatches. The need of precise amounts of Dam methylase may explain the tight and complex control of *dam* gene expression (Løbner-Olesen et al. 1992).

4 Control of Chromosome Replication by DNA Methylation

Initiation of chromosome replication in *E. coli* requires binding of an ATP-bound form of the initiator protein DnaA to the replication origin (*oriC*), followed by separation of the two strands of the DNA double helix and loading of DNA helicase (Mott and Berger 2007). However, binding of DnaA at the *oriC* region is only possible if the GATCs located in the region are methylated; a hemimethylated origin is inactive (Messer et al. 1985). Interestingly, the density of GATC sites in the *oriC* region is roughly tenfold higher than the average in the *E. coli* chromosome (Marinus 1996).

DNA replication is not followed by immediate methylation of the *oriC*. Actually, the GATC sites within *oriC* remain hemimethylated for a substantial fraction of the cell cycle (Messer et al. 1985; Boye et al. 2000). Extension of the hemimethylation period is a consequence of *oriC* sequestration by a protein called SeqA, which binds hemimethylated GATC sites and excludes Dam methylase from the *oriC* in the daughter chromosomes (Lu et al. 1994). As long as SeqA-mediated sequestration and concomitant hemimethylation persist, the *oriC* remains inactive and the start of a new replication cycle is delayed (Boye et al. 2000). SeqA binding to hemimethylated GATC sites is not restricted to the replication origin. Binding of SeqA to hemimethylated GATC sites behind the DNA replication fork may play roles in spatial organization of the nucleoid (Waldminghaus and Skarstad 2009) and in sister chromosome cohesion (Joshi et al. 2013).

Aside from these contributions to cellular welfare, the need of SeqA in *Gammaproteobacteria* may be seen as a burden imposed by the use of DNA hemimethylation as a signal for Dam-dependent mismatch repair. If Dam methylase was

not present throughout the cell cycle, accumulation of MutHLS-induced DNA strand breaks might become lethal (Marinus and Casadesus 2009; Løbner-Olesen et al. 2005). However, the constant presence of Dam methylase makes SeqA necessary to prevent quick methylation of *oriC* and subsequent overinitiation of chromosome replication.

In *Caulobacter crescentus*, the cell cycle is controlled by a complex genetic and epigenetic circuit that includes the CcrM methylase, but the mechanisms involved are more complicated than Dam-mediated control in *Gammaproteobacteria* and remain incompletely understood (Mohapatra et al. 2014). The presence of five GANTC sites in the replication origin of the *Caulobacter* chromosome (*Cori*) turned out to be misleading as these sites are dispensable for chromosome replication control (Gonzalez et al. 2014). However, the involvement of CcrM in cell cycle control is beyond question (Kozdon et al. 2013; Fioravanti et al. 2013; Mohapatra et al. 2014). It is conceivable that CcrM methylation may regulate the *Caulobacter* cell cycle, at least in part, by controlling transcription of genes that encode cell cycle regulators (Fioravanti et al. 2013).

5 Regulation of Bacterial Gene Expression by DNA Methylation

If a DNA methyltransferase target is embedded in a promoter or a regulatory region, its methylation state can modulate binding of RNA polymerase or transcription factors, thus making transcription responsive to DNA methylation. Even though classical examples of transcriptional regulation by DNA methylation involve either Dam or CcrM (Wion and Casadesus 2006; Casadesus and Low 2006; Mohapatra et al. 2014), any DNA methyltransferase can potentially control transcription if it happens to methylate a DNA target at a promoter or at a nearby region involved in transcriptional regulation (Chao et al. 2015; Sanchez-Romero et al. 2015) (Table 2).

Gene expression changes in mutants lacking a DNA methyltransferase do not necessarily indicate that transcription of those genes is DNA methylation sensitive. An example is found in the DNA damage responsive SOS regulon, which shows increased activity in *dam* mutants (Peterson et al. 1985; Torreblanca and Casadesus 1996). However, transcription of the SOS regulon is not controlled by Dam methylation, and activation of SOS functions in *dam* mutants is a consequence of double-strand breakage caused by the MutHLS system in the absence of DNA strand discrimination (Marinus 1996; Marinus and Casadesus 2009). To confirm DNA methylation-dependent transcription, genetic or transcriptomic evidence must be followed by mutational analysis of the putative methyltransferase target(s). If elimination of one or more target sites (e.g., GATC, GANTC, etc.) abolishes DNA methylation-dependent control, one may tentatively conclude that transcription of the gene is DNA methylation-dependent indeed. A potential problem of such tests is that site-directed mutagenesis can alter the binding site of a transcriptional regulator and/or impair promoter function. Verification of a methylation-sensitive DNA-protein interaction requires electrophoretic mobility shift analysis and/or DNAse I

Table 2 Examples of bacterial loci under transcriptional control by N^6-methyladenine

Species	Locus	Methylation-sensitive protein or protein complex	Active state of the promoter or the regulatory region	Function	Reference
E. coli	*tnp(IS10)*	RNA polymerase	Hemimethylated	Transposition	Roberts et al. (1985)
S. enterica	*traJ*	Lrp	Hemimethylated	Plasmid transfer	Camacho and Casadesus (2005)
E. coli	*dnaA*	SeqA	Methylated	DNA replication	Kucherer et al. (1986)
Caulobacter crescentus	*ctrA*	GcrA	Hemimethylated	Cell cycle control	Gonzalez et al. (2014)
Caulobacter crescentus	*ftsZ*	Unknown	Methylated Hemimethylated	Cell division	Gonzalez and Collier (2013)
Caulobacter crescentus	*mipZ*	Unknown	Methylated	Cell division	Gonzalez and Collier (2013)
E. coli	*papBA*	Lrp	Methylation pattern	Fimbriae	Blyn et al. (1990)
E. coli	*agn43*	OxyR	Methylated	Adhesion	Haagmans and van der Woude (2000)
S. enterica	*std*	HdfR	Nonmethylated	Fimbriae	Jakomin et al. (2008)
E. coli	*sciH*	Fur	Nonmethylated	Type VI secretion	Brunet et al. (2011)
S. enterica	*gtr*	OxyR	Methylation pattern	Lipopolysaccharide modification	Broadbent et al. (2010)
S. enterica	*opvAB*	OxyR	Methylation pattern	Lipopolysaccharide modification	Cota et al. (2016)

footprinting using methylated and nonmethylated DNA substrates (Camacho and Casadesus 2002). In certain cases, hemimethylated DNA substrates need to be used also (Camacho and Casadesus 2005).

Bioinformatic prediction of genes whose transcription is controlled by DNA methylation is also intricate. Dam methylation, for instance, can regulate a promoter from distant regulatory sites, sometimes more than 100 base pairs away from the transcription start site. Because the average distance between neighbor GATC sites in the *E. coli* chromosome is 214 base pairs (Hénaut et al. 1996), many promoters contain GATC sites at distances potentially relevant for transcriptional control. Defining the region that needs to be examined for potentially relevant GATCs can thus be difficult. Another limitation of bioinformatic prediction is that the presence of a methylatable site at a seemingly critical position can be misleading. For instance, the *cre* gene of bacteriophage P1 contains two promoters with GATC sites, but Dam methylation controls transcription from one promoter only (Sternberg et al. 1986). On the other hand, genes lacking methylatable targets can be under indirect DNA methylation control if their expression is controlled by a cell factor whose synthesis is DNA methylation dependent. For instance, the cluster of virulence genes known as *Salmonella* pathogenicity island 1 (SPI-1) shows reduced expression in *dam* mutants (Balbontin et al. 2006; Lopez-Garrido and Casadesus 2010). However, regulation of SPI-1 by Dam methylation is indirect and involves StdE and StdF, two proteins encoded by the Dam-dependent *std* operon (Lopez-Garrido and Casadesus 2012).

Studies with the Dam and CcrM model enzymes suggest that DNA methylation-dependent transcriptional controls can be classified into two main classes: (i) clock-like controls that use the methylation state of DNA (methylation or hemimethylation) as a signal to couple gene expression to a specific stage of the cell cycle (Low and Casadesus 2008) and (ii) switch-like controls that turn off and on gene expression, sometimes in a reversible manner, upon formation of DNA methylation patterns. The latter are combinations of methylated and nonmethylated sites reminiscent of the DNA methylation patterns found in eukaryotic chromosomes (Low and Casadesus 2008; Casadesus and Low 2013).

5.1 Temporal Regulation of Gene Expression by DNA Adenine Methylation

RNA polymerase and certain transcription factors can discriminate DNA hemimethylation from DNA methylation in both strands (Wion and Casadesus 2006; Casadesus and Low 2006; Low and Casadesus 2008; Marinus and Casadesus 2009). This discrimination can have physiological significance as DNA hemimethylation is indicative of active growth while two-strand DNA methylation is a hallmark of growth halt. Hemimethylation can either activate or repress gene expression, but activation seems to be more common than repression (Casadesus and Low 2006; Wion and Casadesus 2006). As a rule, nonmethylation of GATC sites is not a physiological state (with remarkable exceptions that will be discussed below).

Nevertheless, genetic and biochemical screens using *dam* mutants have proven useful to identify loci under Dam methylation control, including genes activated by Dam hemimethylation. A tentative explanation for this success is that at certain GATC sites nonmethylation and hemimethylation may provide similar signals, thus producing similar phenotypes (Torreblanca and Casadesus 1996; Oshima et al. 2002; Balbontin et al. 2006).

The insertion element IS10 provides a classical example of transcriptional activation by DNA hemimethylation. The promoter of the IS10 transposase gene contains a GATC site that overlaps the −10 module. Methylation of this GATC prevents transcriptional initiation, presumably by hindrance of RNA polymerase binding (Roberts et al. 1985). When the replication fork passes by the IS10 transposase promoter, the GATC site becomes hemimethylated, and hemimethylation permits transient transcription (Roberts et al. 1985). An additional feature of Dam-dependent control of the IS10 transposase promoter is strand specificity: DNA replication generates two daughter IS10 elements that are identical except for their GATC hemimethylation pattern. However, transcription of the transposase gene is permitted only in the hemimethylated IS10 species that contains N6-methyladenine in the template (noncoding) DNA strand (Roberts et al. 1985).

Another case of transcriptional activation by strand-specific DNA hemimethylation is found in the *traJ* gene of the *Salmonella* virulence plasmid, a relative of the *E. coli* F episome (Camacho and Casadesus 2005). The *traJ* gene encodes a transcription factor, and its expression is controlled by multiple cell factors including Lrp, a global bacterial regulator (Camacho and Casadesus 2002). Lrp activates *traJ* transcription by binding two cognate sites upstream of the *traJ* promoter, one of which contains a GATC (Camacho and Casadesus 2002). Methylation of this GATC impairs Lrp binding and prevents *traJ* transcription (Fig. 2). When replication occurs and the GATC site becomes hemimethylated, Lrp binding activates transcription of *traJ* in one of the daughter plasmid molecules (Camacho and Casadesus 2005). As in the case of IS10, it is noteworthy that two DNA molecules with identical nucleotide sequence can acquire distinct epigenetic properties upon addition of a single methyl group to the template (noncoding) DNA strand.

In both IS10 and *traJ*, activation of transcription by DNA adenine hemimethylation may permit the production of potentially dangerous gene products during active growth only. Strand-specific DNA hemimethylation may further restrain synthesis of such products. A low amount of IS10 transposase may prevent multiple transposition events and/or other transposase-mediated DNA rearrangements (Casadesus and Low 2006; Low and Casadesus 2008). Furthermore, coupling of transposition to DNA replication, a stage of the cell cycle in which two daughter chromosomes exist, may decrease the chances of lethal transposition (Casadesus and Low 2006). In the case of TraJ, plasmid replication may provide a signal of cellular welfare, and Lrp dependence may ensure that the physiological conditions are appropriate to undertake the energy-consuming process of mating. In addition, TraJ synthesis in only one of the daughter plasmids may relieve the burden caused by the synthesis of *tra* operon products and building of the conjugative pilus (Camacho and Casadesus 2002, 2005). Another

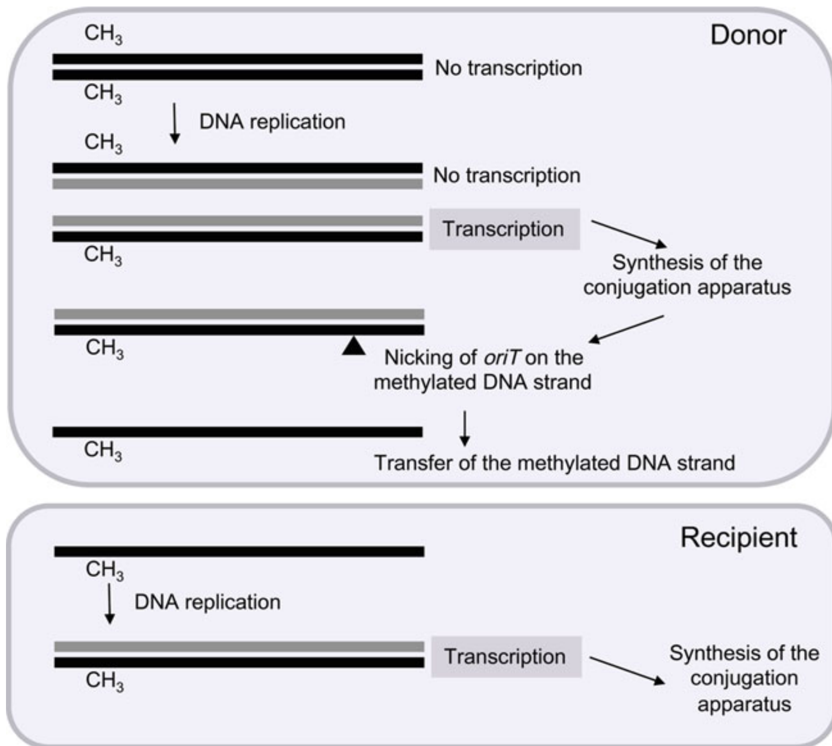


Fig. 2 Activation of *traJ* transcription by strand-specific DNA adenine hemimethylation. Lrp can bind only to the plasmid molecule that carries a methyl group in the noncoding DNA strand. As a consequence, plasmid replication generates two epigenetic states in the *traJ* gene and permits *traJ* transcription in one daughter plasmid molecule only. Reconstruction of the active *traJ* state in the recipient cell after mating is hypothetical

curious aspect of *traJ* regulation is the possibility that the active hemimethylated state can be transmitted to the recipient cell upon plasmid transfer (Fig. 2). If this model is correct, formation of hemimethylated DNA may optimize the spread of the plasmid: as far as recipient cells are available, new donors will be formed by a positive feedback loop. This phenomenon may contribute to explain the old observation that a limiting factor for plasmid spread is the number of recipient cells (Cullum et al. 1978).

In *Alphaproteobacteria*, chromosome hemimethylation is longer-lived than in *Gammaproteobacteria*, especially at loci located near the origin of replication (Collier 2009). While DNA methylation does not seem to play roles in chromosome replication nor in mismatch repair (Gonzalez et al. 2014), hemimethylation of GANTC sites has been shown to activate transcription of *Caulobacter* genes (Gonzalez and Collier 2013; Gonzalez et al. 2014). A relevant example is the cell cycle regulatory gene *ctrA* (Mohapatra et al. 2014; Gonzalez et al. 2014), and others may exist (Mohapatra et al. 2014).

An example of transcriptional repression by DNA adenine hemimethylation is found in the *dnaA* gene of *E. coli* (Marinus 1996). The *dnaA* gene maps near *oriC*, the origin of chromosome replication, and its transcription is driven by three promoters. One of the promoters (*dnaA2*) contains three GATC sites and is only active if they are methylated (Braun and Wright 1986; Kucherer et al. 1985). After DNA replication, the GATC-rich *oriC*-*dnaA* region becomes hemimethylated and is sequestered by SeqA (Lu et al. 1994). Because sequestration prevents Dam methylase activity, the *oriC*-*dnaA* region remains hemimethylated. Initiation of the following chromosome replication round will thus require SeqA release from the *dnaA* promoter and subsequent GATC methylation by the Dam methylase to permit DnaA synthesis. Transcription will be transiently allowed until the next round of DNA replication has occurred (Waldminghaus and Skarstad 2009). Transcriptional repression by hemimethylation has been also described in the cell division genes *ftsZ* and *mipZ* of *Caulobacter* (Gonzalez et al. 2014).

5.2 Regulation of Bacterial Transcription by Formation of DNA Adenine Methylation Patterns

As a rule, hemimethylation of GATC sites in gammaproteobacterial genomes is transient: the Dam methylase trails the DNA replication fork at a relatively short distance, and methylation of the daughter DNA strand restores two-strand GATC methylation (Marinus 1996; Wion and Casadesus 2006). However, the activity of the Dam methylase at specific GATC sites can be hindered by binding of proteins, in a manner reminiscent of sequestration of *oriC* by SeqA (Blyn et al. 1990; Wang and Church 1992). As a consequence, a fraction of GATC sites in the genome of *E. coli* are stably undermethylated (hemimethylated or nonmethylated) (Blyn et al. 1990; Wang and Church 1992). Because active demethylation is not known to occur in bacteria, competition between specific DNA-binding proteins and Dam methylase is the only known mechanism that generates stable undermethylation (Casadesus and Low 2006). Nonmethylation occurs when DNA methylase activity is blocked at least during two consecutive DNA replication rounds. Some undermethylated GATC sites show distinct methylation states depending on growth conditions, suggesting that undermethylation might be the consequence of protein binding in response to physiological or environmental stimuli (Ringquist and Smith 1992; Tavazoie and Church 1998; Hale et al. 1994). Recent findings suggest that undermethylation of GATC sites may not be an exception: orphan DNA methylases other than Dam seem to perform incomplete methylation of their target sites as well (Blow et al. 2016).

Hindrance of Dam methylation by competing proteins requires that the processivity of Dam methylase is reduced. This reduction typically occurs at GATC sites that are part of GATC clusters (two or more GATC sites separated by short distances) and contain AT-rich sequences at their boundaries (Peterson and Reich 2006; Coffin and Reich 2008). Non-processive GATC sites have been found, for instance, in DNA-binding sequences for Lrp, OxyR, Fur, and other transcription

factors, and correlations have been made between DNA methylation states and gene expression patterns (Brunet et al. 2011; Casadesus and Low 2006, 2013; Sanchez-Romero et al. 2015). It must be noted, however, that correlations of this kind are not universal: undermethylated GATC sites that do not seem to control gene expression have also been described (Casadesus and Low 2006; van der Woude et al. 1998).

Formation of DNA methylation patterns has been described in several loci that control programmed, reversible ON-OFF switching of gene expression, a phenomenon known as phase variation (Casadesus and D'Ari 2002; van der Woude 2006, 2011). Phase variation generates bacterial lineages and may facilitate evasion of the host immune system by bacterial pathogens (van der Woude and Bäumler 2004). In other cases, phase variation confers protection against bacteriophage infection (Kim and Ryu 2012; Cota et al. 2015).

A classical example of Dam-dependent control of phase variation is the *pap* operon of uropathogenic *E. coli* strains (Blyn et al. 1990; van der Woude et al. 1996; Braaten et al. 1994). The *pap* operon encodes fimbrial adhesins mediating adherence to the urinary tract epithelium. Populations of uropathogenic *E. coli* contain a mixture of *pap*^{ON} and *pap*^{OFF} cells, and the *pap*^{OFF} subpopulation is always larger because switching is skewed toward the OFF state (van der Woude et al. 1996; Hernday et al. 2002). The *pap*^{ON} and *pap*^{OFF} subpopulations harbor distinct DNA methylation patterns in the *pap* regulatory region, which contains two GATC sites of the reduced processivity type, the proximal site ($GATC_{prox}$), and the distal site ($GATC_{dist}$). In the OFF state, $GATC_{prox}$ is nonmethylated and $GATC_{dist}$ is methylated. In the ON state, $GATC_{prox}$ is methylated and $GATC_{dist}$ is nonmethylated (Fig. 3).

The methylation-blocking protein that creates Dam methylation patterns at the *pap* operon is the global regulator Lrp. The upstream regulatory region (UAS) of *pap* contains six sites for Lrp binding, and two of these sites contain GATCs. When the *pap* operon is not transcribed (OFF state), Lrp is bound to the three downstream sites and represses transcription, probably by preventing RNA polymerase binding. Binding to the downstream sites reduces Lrp affinity for the upstream sites, generating a feedback loop that propagates the OFF state (Hernday et al. 2003). Occupancy of the downstream sites prevents methylation of $GATC_{prox}$, while $GATC_{dist}$ is methylated. This DNA methylation pattern undergoes endless propagation unless a protein called PapI is present (Kaltenbach et al. 1995; Hernday et al. 2003).

Expression of the switching factor PapI is low and probably noisy (Hernday et al. 2002, 2003). Above a critical threshold, PapI stimulates translocation of Lrp to the upstream three binding sites in the *pap* UAS (van der Woude et al. 1996; Casadesus and Low 2006). Binding of Lrp and PapI prevents methylation of $GATC_{dist}$, which becomes nonmethylated. In turn, $GATC_{prox}$ is no longer bound by Lrp and is methylated by the Dam methylase. This configuration ($GATC_{dist}$ nonmethylated, $GATC_{prox}$ methylated) permits *pap* transcription (Casadesus and Low 2006; van der Woude et al. 1996). A positive feedback loop sustains the *pap*^{ON} state: one of the proteins encoded by the *pap* operon, PapB, enhances transcription of the *papI* gene (van der Woude et al. 1996). The ON state is heritable, and under laboratory conditions is perpetuated during 10–12 generations on average, probably with large fluctuations (Casadesus and Low 2006).

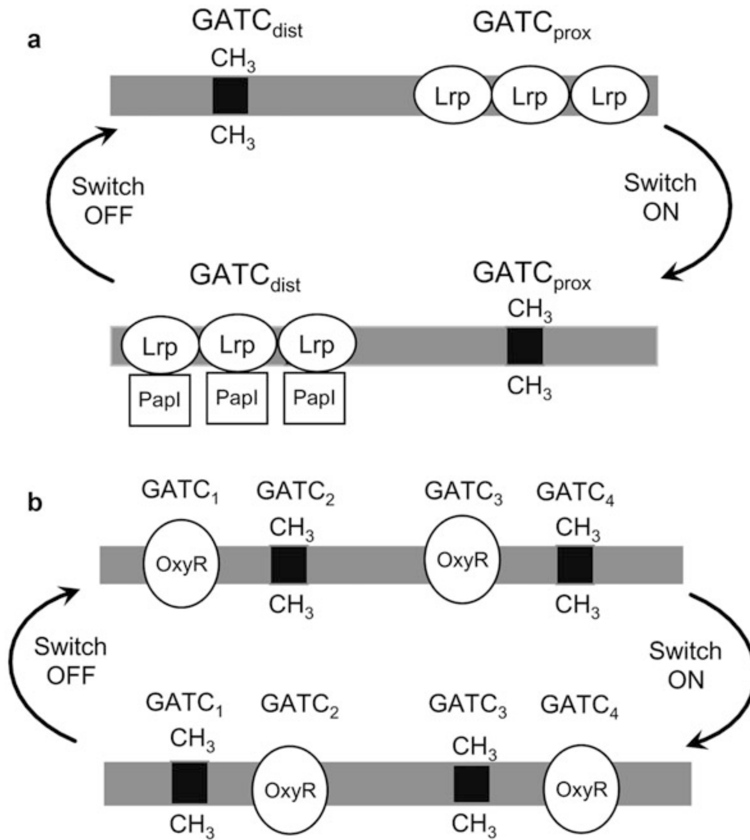


Fig. 3 Phase variation in the *pap* and *opvAB* operons. (**a**) Lrp binding at the downstream sites within the pap regulatory region blocks transcription and prevents methylation of the GATCprox site near the promoter. Transition to phase ON occurs when PapI stimulates translocation of Lrp to the upstream sites, permitting methylation of the GATCprox site and RNA polymerase binding. (**b**) At the opvAB control region, alternative patterns of OxyR binding and GATC methylation are found in OpvABOFF and OpvABON bacterial cell lineages

Switching from ON to OFF requires a decrease in the concentration of PapI below a critical threshold, perhaps by proteolytic degradation (van der Woude et al. 1996). In the absence of PapI, Lrp is unable to bind the upstream regulatory sites and translocates to the downstream sites. Release of the upstream sites by Lrp permits methylation of GATC$_{dist}$, and Lrp translocation to the downstream sites hinders methylation of GATC$_{prox}$ (van der Woude et al. 1996; Casadesus and Low 2006). The *pap*^{OFF} pattern (GATC$_{dist}$ methylated, GATC$_{prox}$ nonmethylated) is thus restored (Casadesus and Low 2006, 2013; Low and Casadesus 2008; van der Woude et al. 1996).

Other phase variation systems regulated by Dam methylation and Lrp are the *foo*, *clp*, and *pef* fimbrial operons, and the architecture of their regulatory regions is

reminiscent of *pap* (Casadesus and Low 2006). Certain phase variation loci controlled by Dam methylation use DNA-binding regulators other than Lrp, including OxyR and HdfR (Table 2). A paradigm of this class is the *E. coli agn43* locus, which encodes an outer membrane protein involved in biofilm formation and host-pathogen interaction (Henderson and Owen 1999; Danese et al. 2000; Luthje and Brauner 2010). Binding of OxyR to the *agn43* regulatory region blocks methylation of three GATC sites and inhibits *agn43* transcription (Waldron et al. 2002; Haagmans and van der Woude 2000). Switching to the *agn43*^{ON} state requires full (two-strand) GATC methylation and may be facilitated by the fact that the *agn43* GATC sites do not have flanking sequences able to reduce the processivity of Dam methylase. Hence, if the GATCs are not bound by OxyR, Dam will processively methylate them. Switching to the *agn43*^{OFF} state can occur after DNA replication, when the three GATCs are hemimethylated (Wallecha et al. 2002). OxyR has a higher affinity for *agn43* DNA containing hemimethylated GATCs than for fully methylated *agn43* DNA. Thus, switching to OFF will be possible if OxyR binds to the regulatory region before Dam methylates the GATC sites (Wallecha et al. 2003; Kaminska and van der Woude 2010).

Other phase variation systems reminiscent of *agn43* include the *gtr* (glycotransferase) locus of bacteriophage P22 (Broadbent et al. 2010), certain *gtr* loci of the *Salmonella* chromosome (Davies et al. 2013), and the *Salmonella opvAB* operon (Cota et al. 2012). All these loci encode proteins that modify the bacterial lipopolysaccharide and are controlled by Dam methylation and OxyR. In the *opvAB* operon, binding of OxyR generates distinct patterns of DNA methylation in OpvAB^{ON} and OpvAB^{OFF} cells, and Dam-dependent regulation is especially complex as it involves two OxyR binding sites and 4 GATC sites and requires the activity of co-regulatory proteins that may induce DNA looping (Cota et al. 2016).

In the last decade, DNA adenine methylation by certain phase-variable type III restriction-modification systems has been found to regulate expression of specific genes, giving rise to a phase-variable regulon or "phasevarion" (Vasu and Nagaraja 2013). Certain phasevarions conserve their restriction-modification activity (Fox et al. 2007b); in others, however, the modification gene (*mod*) remains active but the type III restriction enzyme is inactivated by mutation. Phase-variable synthesis of the Mod methylase generates two subpopulations of bacterial cells, one of which contains N6-methyladenine in the genome while the other subpopulation does not. As a consequence, each lineage shows a distinct pattern of gene expression which affects DNA methylation-sensitive loci (Srikhanta et al. 2005, 2009, 2010). Hence, the Mod enzyme of restriction-deficient phasevarions can be considered a functional analog of solitary DNA adenine methyltransferases like Dam and CcrM (Srikhanta et al. 2010).

A difference between phasevarions and individual phase variation systems such as *pap* and *agn43* is that the cell subpopulations generated by a phasevarion differ in multiple phenotypic traits. For instance, the ModA1 phasevarion of the respiratory pathogen *Haemophilus influenzae* may control at least 15 genes, and the ModA11 phasevarion of *Neisseria meningitidis* may control up to 80 genes (Srikhanta et al. 2010). In the human pathogens *Haemophilus influenzae*, *Neisseria*

meningitidis, *Neisseria gonorrhoeae,* and *Helicobacter pylori*, the loci under Mod control include genes with roles in envelope structure, synthesis of flagella, virulence, and stress responses (Srikhanta et al. 2010). In *N. gonorrhoeae,* lack of the ModA13 methylase alters antimicrobial resistance, invasion of epithelial cells, and biofilm formation (Srikhanta et al. 2009). A phasevarion may also be present in *Moraxella catarrhalis,* generating cell lineages adapted to different human organs (Blakeway et al. 2014).

5.3 Regulation of Bacterial Gene Expression by DNA Cytosine Methylation

The existence of Dcm, a solitary C5-methylcytosine methyltransferase of enteric bacteria, has been a long-lasting paradox (Marinus 1996; Marinus and Casadesus 2009). Hydrolytic deamination of C5-methylcytosine produces thymine, generating T/G mismatches. Enteric bacteria possess a repair system that restores C/G pairs before replication, the so-called very-short-patch (VSP) repair system. Despite the existence of VSP, mutations due to C5-methylcytosine deamination are frequent, especially in the stationary phase (Poole et al. 2001). Mutational hot spots are thus created by formation of C5-methylcytosine, and the potential benefits of C5-methylcytosine formation remain a mystery since loss of Dcm does not have obvious phenotypic consequences, at least under laboratory conditions (Marinus and Casadesus 2009).

Recent studies, however, suggest that DNA cytosine methylation may play physiological roles in *E. coli*, perhaps of a subtle nature, including regulation of gene expression. *E. coli dcm* mutants show increased expression of the sigma factor RpoS (Kahramanoglou et al. 2012) and overexpress a membrane protein involved in ethidium bromide transport (Militello et al. 2013). In *Helicobacter pylori*, lack of an orphan C5-methylcytosine methyltransferase known as HpyAVIBM alters the expression of genes involved in motility, adhesion, and virulence (Kumar et al. 2012). Because DNA repeats are present in the *hpyAVIBM* coding sequence, it is conceivable that repeat expansion and/or contraction might cause phase-variable expression, thus forming a C5-cytosine phasevarion (Kumar et al. 2012). Equally or more exciting is the observation that a DNA C5-methylcytosine methyltransferase acquired by horizontal transfer can become integrated into the genetic networks of the cell in a rapid or sudden manner (Chao et al. 2015), thus promoting saltational evolutionary change.

6 Bacterial Methylomes

The difficulty to detect methylated bases (particularly N6-methyladenine, the most common modification found in bacterial genomes) has slowed down for decades the study of DNA methylation in bacteria. In the last few years, however, advances in nucleic acid sequencing technology have permitted the development of

procedures that detect DNA base modifications as an integral part of the sequencing method. For instance, single-molecule real-time (SMRT) sequencing monitors the activity of a single DNA polymerase that uses fluorescent nucleotides to synthesize DNA complementary to a template (Flusberg et al., 2010; Davis et al., 2013). Addition of a nucleotide is detected as a pulse of fluorescence whose color identifies the nucleotide. The sequencing device monitors not only the fluorescence pulse associated with each incorporated nucleotide but also the time between successive pulses (interpulse duration (IPD)). The IPD is statistically longer if the template contains a methylated base, and the kinetic signatures of N6-methyladenine and N4-methylcytosine templates can be distinguished (Flusberg et al., 2010). At a given position, an altered IPD ratio between native (methylated) DNA and PCR-amplified (nonmethylated) DNA identifies a methylated nucleotide in the template DNA (Flusberg et al. 2010; Davis et al. 2013). The complete methylation pattern (the "methylome") can thus be obtained, together with the DNA sequence of any DNA molecule (e.g., of a bacterial genome).

The benefits of SMRT sequencing have been immediate. One relevant observation, for instance, is the detection of DNA methylation in many bacterial and archaebacterial genomes (Blow et al. 2016). Equally or more interesting is the fact that a significant fraction of the responsible DNA methyltransferases appear to belong to the orphan type, suggesting an evolutionary origin unrelated to protection from restriction (Blow et al. 2016). Methylome analysis also permits to address a variety of biological questions for which traditional methodologies fell short. For instance, inference of the specificity of a DNA methyltransferase belonging to a restriction-modification system was possible only for type II enzymes, based on patterns of DNA cleavage. In contrast, methylome analysis reveals at once all the methylated motifs present in the genome (Murray et al. 2012), an information especially useful in species that harbor multiple restriction-modification systems like *Helicobacter pylori* (Krebes et al. 2014; Lee et al. 2015). Other relevant examples of biological questions amenable to methylome analysis include the physiological consequences of DNA methylase acquisition (Chao et al. 2015) or DNA methylase specificity alteration (Furuta et al. 2014), the contribution of prophage DNA methylases to the physiology of the bacterial host (Fang et al. 2012), the polymorphism in DNA methylase assortment within a species (Pirone-Davies et al. 2015), and the occurrence of DNA methylation in small bacterial genomes (Lluch-Senar et al. 2013).

A limitation of SMRT sequencing is that IPD values are calculated as averages from multiple assessments at each genomic position. Therefore, the method is unable to discriminate potential variations in DNA methylation patterns within a bacterial population. This limitation is serious if one considers, for instance, that phase variation generates subpopulations with distinct DNA methylation patterns in gene control regions (Blyn et al. 1990; Correnti et al. 2002; Broadbent et al. 2010; Cota et al. 2012). As an example, the identification of the Dam methylation patterns associated with *ovpAB* phase variation required independent examination of $OpvAB^{ON}$ and $OpvAB^{OFF}$ bacterial lineages (Cota et al. 2016). The same limitation may apply to the identification of methylation patterns produced by phase-variable

Mod enzymes. The problem can be expected to be solved by a novel procedure, known as SMALR (single-molecule modification analysis of long reads), that can detect heterogeneous DNA methylation patterns in mixed populations of bacteria (Beaulaurier et al. 2015).

7 Roles of DNA Methylation in Bacterial Pathogenesis

Evidence for a relationship between Dam methylation and bacterial virulence was initially provided by the regulation of adhesin-encoding genes like the *pap* operon of *E. coli* (Blyn et al. 1990). However, the role of Dam methylation in the infection of model animals was first investigated in *Salmonella enterica*. A simple genetic approach was to compare the median lethal dose (LD_{50}) of a Dam methylase mutant with that of the wild type upon infection of mice. Additional details about the infection process were provided by examination of animal organs and in vitro studies using cell cultures. Lack of Dam methylation was found to cause severe attenuation in the mouse model of typhoid: the LD_{50} of a *Salmonella dam* mutant is 10,000-fold higher than that of the wild type by the oral route and 1,000-fold higher intraperitoneally (Heithoff et al. 1999; Garcia-Del Portillo et al. 1999). This extreme attenuation seems to be caused by a combination of defects: (i) reduced capacity to interact with the intestinal epithelium, due to inefficient activation of genes in pathogenicity island I (SPI-1) (Balbontin et al. 2006; Lopez-Garrido and Casadesus 2010) and impaired secretion of effectors (Giacomodonato et al. 2009b); (ii) reduced motility, due to chaotic expression of flagellar and chemotaxis genes (Balbontin et al. 2006); (iii) envelope instability, with release of outer membrane vesicles and leakage of proteins (Pucciarelli et al. 2002); (iv) sensitivity to bile salts, a defect that may compromise survival in the hepatobiliary tract (Pucciarelli et al. 2002; Heithoff et al. 2001); and (v) deficient biofilm formation (Aya Castaneda Mdel et al. 2015).

Virulence-related defects associated with loss of DNA methylation have been reported in other pathogens as well (Heusipp et al. 2007; Marinus and Casadesus 2009; Giacomodonato et al. 2009a). In *Streptococcus mutans,* Dam methylation may control genes involved in dental cariogenesis (Banas et al. 2011). In certain strains of *Haemophilus influenzae*, Dam methylation is necessary for invasion of endothelial and epithelial cell lines (Watson et al. 2004). Reduced invasion of epithelial cells by *dam* mutants is likewise observed in the periodontal disease agent *Aggregatibacter* (previously *Actinobacillus) actinomycetemcomitans* (Chen et al. 2003) and perhaps in the intestinal pathogen *Campylobacter jejuni* (Kim et al. 2008).

An intriguing connection between DNA adenine methylation and virulence is found in the gram-positive pathogen *Mycobacterium tuberculosis*. Strains of the Euro-American lineage of *M. tuberculosis* harbor a DNA adenine methyltransferase called MamA, which is absent from strains of the Beijing lineage. MamA methylation appears to control survival in hypoxia, a stress condition found during human infection, and may regulate expression of a number of *M. tuberculosis* genes (Shell

et al. 2013). A different DNA adenine methyltransferase is found in the Beijing lineage, suggesting that strain-specific differences in DNA methylation may control lineage-specific features (Shell et al. 2013).

In bacterial species where the viability of DNA methylase mutants is impaired, the effects of DNA methylase overproduction can be examined. Among *Gammaproteobacteria*, overproduction of Dam methylase in *Yersinia enterocolitica* increases invasion, probably by altered synthesis of invasin, a protein that stimulates phagocytosis, and by changes in the composition of O antigen in the lipopolysaccharide (Fälker et al. 2007). In addition, virulence-related defects of *Y. enterocolitica* Dam-overproducing strains include enhanced motility and impaired secretion of the pathogenicity factors called "*Yersinia* outer proteins" (Fälker et al. 2007). Among *Alphaproteobacteria*, overproduction of the CcrM methylase reduces *Brucella abortus* proliferation inside macrophages, suggesting that CcrM methylation may control intracellular replication, which is a hallmark of brucellosis (Robertson et al. 2000).

The involvement of DNA adenine methylation in the virulence of both alpha- and gammaproteobacterial pathogens has raised the possibility of using DNA methylase mutants as live vaccines (Heithoff et al. 2001; Giacomodonato et al. 2004). Such vaccines have been assayed indeed against *S. enterica*, *Haemophilus influenzae*, and *Yersinia pseudotuberculosis* (Marinus and Casadesus 2009; Mohler et al. 2012). Hypermutability, however, is a problematic trait and may become a hurdle to use *dam* mutants of *Gammaproteobacteria* as live vaccines, at least in humans.

The relationship between DNA methylation and bacterial virulence has also fostered the search for DNA methylase inhibitors that might serve as antibacterial drugs (Mashhoon et al. 2004, 2006; Benkovic et al. 2005). Because adenine methylation is rare or absent in mammalian cells (Ratel et al. 2006), inhibitors of Dam-like or CcrM-like methyltransferases can be expected to be harmless for the host. In the case of Dam methylation, its dispensable nature in most *Gammaproteobacteria* and its absence in other bacterial taxa (Løbner-Olesen et al. 2005) should make inhibitors harmless for the normal microbiota. In pathogens, however, Dam methylase inhibitors can be expected to attenuate virulence by transforming wild-type bacteria into phenocopies of *dam* mutants. Such drugs might have broad spectrum as Dam and CcrM methylation seem to control virulence in a variety of bacterial pathogens (Marinus and Casadesus 2009; Sanchez-Romero et al. 2015). Again, a negative aspect is that inhibition of DNA adenine methylation can be expected to increase the mutation rate in bacterial species that use Dam-dependent strand discrimination for DNA mismatch repair. Unfortunately, such species include *E. coli* intestinal commensals and perhaps other members of the human microbiota.

Acknowledgments I thank Ignacio Cota, María A. Sánchez-Romero, and Lucía García-Pastor for discussions and Modesto Carballo, Laura Navarro, and Cristina Reyes (CITIUS) for assistance. Research in my laboratory is supported by grants BIO2013-44220-R from the Ministerio de Economía y Competitividad of Spain (MINECO) and the European Regional Fund, PCIN-2015-131 by MINECO and Infect-ERA, and CVI-5879 from the Consejería de Innovación, Ciencia y Empresa, Junta de Andalucía, Spain.

References

Albu RF, Jurkowski TP, Jeltsch A. The *Caulobacter crescentus* DNA-(adenine N6)-methyltransferase CcrM methylates DNA in a distributive manner. Nucleic Acids Res. 2011;40(4):1708–16.

Arber W. Genetic variation: molecular mechanisms and impact on microbial evolution. FEMS Microbiol Rev. 2000;24(1):1–7.

Arber W, Linn S. DNA modification and restriction. Annu Rev Biochem. 1969;38:467–500.

Aya Castaneda Mdel R, Sarnacki SH, Noto Llana M, Lopez Guerra AG, Giacomodonato MN, Cerquetti MC. Dam methylation is required for efficient biofilm production in *Salmonella enterica* serovar Enteritidis. Int J Food Microbiol. 2015;193:15–22.

Balbontin R, Rowley G, Pucciarelli MG, Lopez-Garrido J, Wormstone Y, Lucchini S, Garcia-Del Portillo F, Hinton JC, Casadesus J. DNA adenine methylation regulates virulence gene expression in *Salmonella enterica* serovar Typhimurium. J Bacteriol. 2006;188(23):8160–8.

Banas JA, Biswas S, Zhu M. Effects of DNA methylation on expression of virulence genes in *Streptococcus mutans*. Appl Environ Microbiol. 2011;77(20):7236–42.

Beaulaurier J, Zhang XS, Zhu S, Sebra R, Rosenbluh C, Deikus G, Shen N, Munera D, Waldor MK, Chess A, Blaser MJ, Schadt EE, Fang G. Single molecule-level detection and long read-based phasing of epigenetic variations in bacterial methylomes. Nat Commun. 2015;6:7438.

Benkovic SJ, Baker SJ, Alley MR, Woo YH, Zhang YK, Akama T, Mao W, Baboval J, Rajagopalan PT, Wall M, Kahng LS, Tavassoli A, Shapiro L. Identification of borinic esters as inhibitors of bacterial cell growth and bacterial methyltransferases, CcrM and MenH. J Med Chem. 2005;48(23):7468–76.

Bergerat A, Guschlbauer W, Fazakerley GV. Allosteric and catalytic binding of S-adenosylmethionine to *Escherichia coli* DNA adenine methyltransferase monitored by 3H NMR. Proc Natl Acad Sci U S A. 1991;88(15):6394–7.

Bertani G, Weigle JJ. Host controlled variation in bacterial viruses. J Bacteriol. 1953;65(2):113–21.

Bhaya D, Davison M, Barrangou R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annu Rev Genet. 2012;45:273–97.

Blakeway LV, Power PM, Jen FE, Worboys SR, Boitano M, Clark TA, Korlach J, Bakaletz LO, Jennings MP, Peak IR, Seib KL. ModM DNA methyltransferase methylome analysis reveals a potential role for *Moraxella catarrhalis* phasevarions in otitis media. FASEB J. 2014;28(12):5197–207.

Blow MJ, Clark TA, Daum CG, Deutschbauer AM, Fomenkov A, Fries R, Froula J, Kang DD, Malmstrom RR, Morgan RD, Posfai J, Singh K, Visel A, Wetmore K, Zhao Z, Rubin EM, Korlach J, Pennacchio LA, Roberts RJ. The epigenomic landscape of prokaryotes. PLoS Genet. 2016;12(2):e1005854.

Blyn LB, Braaten BA, Low DA. Regulation of pap pilin phase variation by a mechanism involving differential *dam* methylation states. EMBO J. 1990;9(12):4045–54.

Boye E, Løbner-Olesen A, Skarstad K. Limiting DNA replication to once and only once. EMBO Rep. 2000;1(6):479–83.

Braaten BA, Nou X, Kaltenbach LS, Low DA. Methylation patterns in *pap* regulatory DNA control pyelonephritis-associated pili phase variation in *E. coli*. Cell. 1994;76(3):577–88.

Braun RE, Wright A. DNA methylation differentially enhances the expression of one of the two *E. coli dnaA* promoters *in vivo* and *in vitro*. Mol Gen Genet. 1986;202(2):246–50.

Broadbent SE, Balbontin R, Casadesus J, Marinus MG, van der Woude M. YhdJ, a nonessential CcrM-like DNA methyltransferase of *Escherichia coli* and *Salmonella enterica*. J Bacteriol. 2007;189(11):4325–7.

Broadbent SE, Davies MR, van der Woude MW. Phase variation controls expression of *Salmonella* lipopolysaccharide modification genes by a DNA methylation-dependent mechanism. Mol Microbiol. 2010;77(2):337–53.

Brunet YR, Bernard CS, Gavioli M, Lloubes R, Cascales E. An epigenetic switch involving overlapping Fur and DNA methylation optimizes expression of a type VI secretion gene cluster. PLoS Genet. 2011;7(7):e1002205.

Calmann MA, Marinus MG. Regulated expression of the *Escherichia coli dam* gene. J Bacteriol. 2003;185(16):5012–4.

Camacho EM, Casadesus J. Conjugal transfer of the virulence plasmid of *Salmonella enterica* is regulated by the leucine-responsive regulatory protein and DNA adenine methylation. Mol Microbiol. 2002;44(6):1589–98.

Camacho EM, Casadesus J. Regulation of *traJ* transcription in the *Salmonella* virulence plasmid by strand-specific DNA adenine hemimethylation. Mol Microbiol. 2005;57(6):1700–18.

Casadesus J, D'Ari R. Memory in bacteria and phage. Bioessays. 2002;24(6):512–8.

Casadesus J, Low D. Epigenetic gene regulation in the bacterial world. Microbiol Mol Biol Rev. 2006;70(3):830–56.

Casadesus J, Low DA. Programmed heterogeneity: epigenetic mechanisms in bacteria. J Biol Chem. 2013;288(20):13929–35.

Chao MC, Zhu S, Kimura S, Davis BM, Schadt EE, Fang G, Waldor MK. A cytosine methyltransferase modulates the cell envelope stress response in the cholera pathogen. PLoS Genet. 2015;11(11):e1005666.

Chen L, Paulsen DB, Scruggs DW, Banes MM, Reeks BY, Lawrence ML. Alteration of DNA adenine methylase (Dam) activity in *Pasteurella multocida* causes increased spontaneous mutation frequency and attenuation in mice. Microbiology. 2003;149:2283–90.

Cheng X. Structure and function of DNA methyltransferases. Annu Rev Biophys Biomol Struct. 1995;24:293–318.

Coffin SR, Reich NO. Modulation of *Escherichia coli* DNA methyltransferase activity by biologically derived GATC-flanking sequences. J Biol Chem. 2008;283(29):20106–16.

Collier J. Epigenetic regulation of the bacterial cell cycle. Curr Opin Microbiol. 2009;12(6):722–9.

Correnti J, Munster V, Chan T, Woude M. Dam-dependent phase variation of Ag43 in *Escherichia coli* is altered in a *seqA* mutant. Mol Microbiol. 2002;44(2):521–32.

Cota I, Blanc-Potard AB, Casadesus J. *STM2209-STM2208 (opvAB)*: a phase variation locus of *Salmonella enterica* involved in control of O-antigen chain length. PLoS One. 2012;7(5):e36863.

Cota I, Bunk B, Sproer C, Overmann J, Konig C, Casadesus J. OxyR-dependent formation of DNA methylation patterns in OpvAB^{OFF} and OpvAB^{ON} cell lineages of *Salmonella enterica*. Nucleic Acids Res. 2016;44(8):3595–609. doi:10.1093/nar/gkv1483.

Cota I, Sanchez-Romero MA, Hernandez SB, Pucciarelli MG, Garcia-Del Portillo F, Casadesus J. Epigenetic control of *Salmonella enterica* O-antigen chain length: A tradeoff between virulence and bacteriophage resistance. PLoS Genet. 2015;11(11):e1005667.

Cullum J, Collins JF, Broda P. The spread of plasmids in model populations of *Escherichia coli* K12. Plasmid. 1978;1(4):545–56.

Danese PN, Pratt LA, Dove SL, Kolter R. The outer membrane protein, antigen 43, mediates cell-to-cell interactions within *Escherichia coli* biofilms. Mol Microbiol. 2000;37(2):424–32.

Davies MR, Broadbent SE, Harris SR, Thomson NR, van der Woude MW. Horizontally acquired glycosyltransferase operons drive salmonellae lipopolysaccharide diversity. PLoS Genet. 2013;9(6):e1003568.

Davis BM, Chao MC, Waldor MK. Entering the era of bacterial epigenomics with single molecule real time DNA sequencing. Curr Opin Microbiol. 2013;16(2):192–8.

Fälker S, Schilling J, Schmidt MA, Heusipp G. Overproduction of DNA adenine methyltransferase alters motility, invasion, and the lipopolysaccharide O-antigen composition of *Yersinia enterocolitica*. Infect Immun. 2007;75(10):4990–7.

Fang G, Munera D, Friedman DI, Mandlik A, Chao MC, Banerjee O, Feng Z, Losic B, Mahajan MC, Jabado OJ, Deikus G, Clark TA, Luong K, Murray IA, Davis BM, Keren-Paz A, Chess A, Roberts RJ, Korlach J, Turner SW, Kumar V, Waldor MK, Schadt EE. Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. Nat Biotechnol. 2012;30(12):1232–9.

Fioravanti A, Fumeaux C, Mohapatra SS, Bompard C, Brilli M, Frandi A, Castric V, Villeret V, Viollier PH, Biondi EG. DNA binding of the cell cycle transcriptional regulator GcrA depends on N6-adenosine methylation in *Caulobacter crescentus* and other Alphaproteobacteria. PLoS Genet. 2013;9(5):e1003541.

Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SW. Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods. 2010;7(6):461–5.

Fox KL, Dowideit SJ, Erwin AL, Srikhanta YN, Smith AL, Jennings MP. *Haemophilus influenzae* phasevarions have evolved from type III DNA restriction systems into epigenetic regulators of gene expression. Nucleic Acids Res. 2007a;35(15):5242–52.

Fox KL, Srikhanta YN, Jennings MP. Phase variable type III restriction-modification systems of host-adapted bacterial pathogens. Mol Microbiol. 2007b;65(6):1375–9.

Furuta Y, Namba-Fukuyo H, Shibata TF, Nishiyama T, Shigenobu S, Suzuki Y, Sugano S, Hasebe M, Kobayashi I. Methylome diversification through changes in DNA methyltransferase sequence specificity. PLoS Genet. 2014;10(4):e1004272.

Garcia-Del Portillo F, Pucciarelli MG, Casadesus J. DNA adenine methylase mutants of *Salmonella typhimurium* show defects in protein secretion, cell invasion, and M cell cytotoxicity. Proc Natl Acad Sci U S A. 1999;96(20):11578–83.

Giacomodonato MN, Sarnacki SH, Caccuri RL, Sordelli DO, Cerquetti MC. Host response to a *dam* mutant of *Salmonella enterica* serovar Enteritidis with a temperature-sensitive phenotype. Infect Immun. 2004;72(9):5498–501.

Giacomodonato MN, Sarnacki SH, Llana MN, Cerquetti MC. Dam and its role in pathogenicity of *Salmonella enterica*. J Infect Dev Ctries. 2009a;3(7):484–90.

Giacomodonato MN, Sarnacki SH, Llana MN, Garcia Cattaneo AS, Uzzau S, Rubino S, Cerquetti MC. Impaired synthesis and secretion of SopA in *Salmonella* Typhimurium *dam* mutants. FEMS Microbiol Lett. 2009b;292(1):71–7.

Glickman B, van den Elsen P, Radman M. Induced mutagenesis in *dam*- mutants of *Escherichia coli*: a role for 6-methyladenine residues in mutation avoidance. Mol Gen Genet. 1978; 163(3):307–12.

Gonzalez D, Collier J. DNA methylation by CcrM activates the transcription of two genes required for the division of *Caulobacter crescentus*. Mol Microbiol. 2013;88(1):203–18.

Gonzalez D, Kozdon JB, McAdams HH, Shapiro L, Collier J. The functions of DNA methylation by CcrM in *Caulobacter crescentus*: a global approach. Nucleic Acids Res. 2014;42(6):3720–35.

Haagmans W, van der Woude M. Phase variation of Ag43 in *Escherichia coli*: Dam-dependent methylation abrogates OxyR binding and OxyR-mediated repression of transcription. Mol Microbiol. 2000;35(4):877–87.

Hale WB, van der Woude MW, Low DA. Analysis of nonmethylated GATC sites in the *Escherichia coli* chromosome and identification of sites that are differentially methylated in response to environmental stimuli. J Bacteriol. 1994;176(11):3438–41.

Heithoff DM, Enioutina EI, Daynes RA, Sinsheimer RL, Low DA, Mahan MJ. *Salmonella* DNA adenine methylase mutants confer cross-protective immunity. Infect Immun. 2001;69: 6725–30.

Heithoff DM, Sinsheimer RL, Low DA, Mahan MJ. An essential role for DNA adenine methylation in bacterial virulence. Science. 1999;284(5416):967–70.

Hénaut A, Rouxel T, Gleizes A, Moszer I, Danchin A. Uneven distribution of GATC motifs in the *Escherichia coli* chromosome, its plasmids and its phages. J Mol Biol. 1996;257(3):574–85.

Henderson IR, Owen P. The major phase-variable outer membrane protein of *Escherichia coli* structurally resembles the immunoglobulin A1 protease class of exported protein and is regulated by a novel mechanism involving Dam and OxyR. J Bacteriol. 1999;181(7):2132–41.

Herman GE, Modrich P. *Escherichia coli dam* methylase. Physical and catalytic properties of the homogeneous enzyme. J Biol Chem. 1982;257(5):2605–12.

Hernday A, Krabbe M, Braaten B, Low D. Self-perpetuating epigenetic pili switches in bacteria. Proc Natl Acad Sci U S A. 2002;99:16470–6.

Hernday AD, Braaten BA, Low DA. The mechanism by which DNA adenine methylase and PapI activate the pap epigenetic switch. Mol Cell. 2003;12(4):947–57.

Heusipp G, Fälker S, Schmidt MA. DNA adenine methylation and bacterial pathogenesis. Int J Med Microbiol. 2007;297(1):1–7.

Iyer RR, Pluciennik A, Burdett V, Modrich PL. DNA mismatch repair: functions and mechanisms. Chem Rev. 2006;106(2):302–23.

Jakomin M, Chessa D, Baumler AJ, Casadesus J. Regulation of the *Salmonella enterica std* fimbrial operon by DNA adenine methylation, SeqA, and HdfR. J Bacteriol. 2008;190(22): 7406–13.

Jeltsch A. Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. Chembiochem. 2002;3(4):274–93.

Jeltsch A. Maintenance of species identity and controlling speciation of bacteria: a new function for restriction/modification systems? Gene. 2003;317(1–2):13–6.

Joshi MC, Magnan D, Montminy TP, Lies M, Stepankiw N, Bates D. Regulation of sister chromosome cohesion by the replication fork tracking protein SeqA. PLoS Genet. 2013;9(8):e1003673.

Kahng LS, Shapiro L. The CcrM DNA methyltransferase of *Agrobacterium tumefaciens* is essential, and its activity is cell-cycle regulated. J Bacteriol. 2001;183(10):3065–75.

Kahramanoglou C, Prieto AI, Khedkar S, Haase B, Gupta A, Benes V, Fraser GM, Luscombe NM, Seshasayee AS. Genomics of DNA cytosine methylation in *Escherichia coli* reveals its role in stationary phase transcription. Nat Commun. 2012;3:886.

Kaltenbach LS, Braaten BA, Low DA. Specific binding of PapI to Lrp-pap DNA complexes. J Bacteriol. 1995;177(22):6449–55.

Kaminska R, van der Woude MW. Establishing and maintaining sequestration of Dam target sites for phase variation of *agN43* in *Escherichia coli*. J Bacteriol. 2010;192(7):1937–45.

Kim JS, Li J, Barnes IH, Baltzegar DA, Pajaniappan M, Cullen TW, Trent MS, Burns CM, Thompson SA. Role of the *Campylobacter jejuni* Cj1461 DNA methyltransferase in regulating virulence characteristics. J Bacteriol. 2008;190(19):6524–9.

Kim M, Ryu S. Spontaneous and transient defence against bacteriophage by phase-variable glucosylation of O-antigen in *Salmonella enterica* serovar Typhimurium. Mol Microbiol. 2012;86(2):411–25.

Kozdon JB, Melfi MD, Luong K, Clark TA, Boitano M, Wang S, Zhou B, Gonzalez D, Collier J, Turner SW, Korlach J, Shapiro L, McAdams HH. Global methylation state at base-pair resolution of the *Caulobacter* genome throughout the cell cycle. Proc Natl Acad Sci U S A. 2013;110(48):E4658–67.

Krebes J, Morgan RD, Bunk B, Sproer C, Luong K, Parusel R, Anton BP, Konig C, Josenhans C, Overmann J, Roberts RJ, Korlach J, Suerbaum S. The complex methylome of the human gastric pathogen *Helicobacter pylori*. Nucleic Acids Res. 2014;42(4):2415–32.

Kucherer C, Lother H, Kolling R, Schauzu MA, Messer W. Regulation of transcription of the chromosomal *dnaA* gene of *Escherichia coli*. Mol Gen Genet. 1985;205(1):115–21.

Kumar R, Mukhopadhyay AK, Ghosh P, Rao DN. Comparative transcriptomics of *H. pylori* strains AM5, SS1 and their hpyAVIBM deletion mutants: possible roles of cytosine methylation. PLoS One. 2012;7(8):e42303.

Lee WC, Anton BP, Wang S, Baybayan P, Singh S, Ashby M, Chua EG, Tay CY, Thirriot F, Loke MF, Goh KL, Marshall BJ, Roberts RJ, Vadivelu J. The complete methylome of *Helicobacter pylori* UM032. BMC Genomics. 2015;16:424.

Lluch-Senar M, Luong K, Llorens-Rico V, Delgado J, Fang G, Spittle K, Clark TA, Schadt E, Turner SW, Korlach J, Serrano L. Comprehensive methylome characterization of *Mycoplasma genitalium* and *Mycoplasma pneumoniae* at single-base resolution. PLoS Genet. 2013; 9(1):e1003191.

Løbner-Olesen A, Boye E, Marinus MG. Expression of the *Escherichia coli dam* gene. Mol Microbiol. 1992;6(13):1841–51.

Løbner-Olesen A, Skovgaard O, Marinus MG. Dam methylation: coordinating cellular processes. Curr Op Microbiol. 2005;8:154–60.

Loenen WA, Dryden DT, Raleigh EA, Wilson GG, Murray NE. Highlights of the DNA cutters: a short history of the restriction enzymes. Nucleic Acids Res. 2014;42(1):3–19.

Loenen WA, Raleigh EA. The other face of restriction: modification-dependent enzymes. Nucleic Acids Res. 2014;42(1):56–69.

Lopez-Garrido J, Casadesus J. Regulation of *Salmonella enterica* pathogenicity island 1 by DNA adenine methylation. Genetics. 2010;184(3):637–49.

Lopez-Garrido J, Casadesus J. Crosstalk between virulence loci: regulation of *Salmonella enterica* pathogenicity island 1 (SPI-1) by products of the *std* fimbrial operon. PLoS One. 2012;7(1):e30499.

Low DA, Casadesus J. Clocks and switches: bacterial gene regulation by DNA adenine methylation. Curr Opin Microbiol. 2008;11(2):106–12.

Lu M, Campbell JL, Boye E, Kleckner N. SeqA: a negative modulator of replication initiation in *E. coli*. Cell. 1994;77(3):413–26.

Luthje P, Brauner A. Ag43 promotes persistence of uropathogenic *Escherichia coli* isolates in the urinary tract. J Clin Microbiol. 2010;48(6):2316–7.

Marinus MG. Methylation of DNA. In: Neidhardt FC et al. editors. *Escherichia coli* and *Salmonella*. Cellular and molecular biology. 2nd ed. Washington, DC: ASM Press; 1996. p. 782–91.

Marinus MG, Casadesus J. Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more. FEMS Microbiol Rev. 2009;33(3):488–503.

Marinus MG, Morris NR. Biological function for 6-methyladenine residues in the DNA of *Escherichia coli* K12. J Mol Biol. 1974;85(2):309–22.

Marinus MG, Poteete A, Arraj JA. Correlation of DNA adenine methylase activity with spontaneous mutability in *Escherichia coli K-12*. Gene. 1984;28(1):123–5.

Mashhoon N, Carroll M, Pruss C, Eberhard J, Ishikawa S, Estabrook RA, Reich N. Functional characterization of *Escherichia coli* DNA adenine methyltransferase, a novel target for antibiotics. J Biol Chem. 2004;259:52075–81.

Mashhoon N, Pruss C, Carroll M, Johnson PH, Reich N. Selective inhibitors of bacterial DNA adenine methyltransferases. J Biomol Screen. 2006;11:497–506.

Messer W, Bellekes U, Lother H. Effect of *dam* methylation on the activity of the *E. coli* replication origin, *oriC*. EMBO J. 1985;4(5):1327–32.

Militello KT, Mandarano AH, Varechtchouk O, Simon RD. Cytosine DNA methylation influences drug resistance in *Escherichia coli* through increased *sugE* expression. FEMS Microbiol Lett. 2013;350(1):100–6.

Mohapatra SS, Fioravanti A, Biondi EG. DNA methylation in *Caulobacter* and other Alphaproteobacteria during cell cycle progression. Trends Microbiol. 2014;22:528–35.

Mohler VL, Heithoff DM, Mahan MJ, Hornitzky MA, Thomson PC, House JK. Development of a novel in-water vaccination protocol for DNA adenine methylase deficient *Salmonella enterica* serovar Typhimurium vaccine in adult sheep. Vaccine. 2012;30(8):1481–91.

Mott ML, Berger JM. DNA replication initiation: mechanisms and regulation in bacteria. Nat Rev Microbiol. 2007;5(5):343–54.

Murray IA, Clark TA, Morgan RD, Boitano M, Anton BP, Luong K, Fomenkov A, Turner SW, Korlach J, Roberts RJ. The methylomes of six bacteria. Nucleic Acids Res. 2012;40(22):11450–62.

Naito T, Kusano K, Kobayashi I. Selfish behavior of restriction-modification systems. Science. 1995;267(5199):897–9.

Oshima T, Wada C, Kawagoe Y, Ara T, Maeda M, Masuda Y, Hiraga S, Mori H. Genome-wide analysis of deoxyadenosine methyltransferase-mediated control of gene expression in *Escherichia coli*. Mol Microbiol. 2002;45(3):673–95.

Peterson KR, Wertman KF, Mount DW, Marinus MG. Viability of *Escherichia coli* K-12 DNA adenine methylase (*dam*) mutants requires increased expression of specific genes in the SOS regulon. Mol Gen Genet. 1985;201(1):14–9.

Peterson SN, Reich NO. GATC flanking sequences regulate Dam activity: evidence for how Dam specificity may influence *pap* expression. J Mol Biol. 2006;355(3):459–72.

Pirone-Davies C, Hoffmann M, Roberts RJ, Muruvanda T, Timme RE, Strain E, Luo Y, Payne J, Luong K, Song Y, Tsai YC, Boitano M, Clark TA, Korlach J, Evans PS, Allard MW. Genome-

wide methylation patterns in *Salmonella enterica* subsp. *enterica* serovars. PLoS One. 2015;10(4):e0123639.
Poole A, Penny D, Sjoberg BM. Confounded cytosine! Tinkering and the evolution of DNA. Nat Rev Mol Cell Biol. 2001;2(2):147–51.
Pucciarelli MG, Prieto AI, Casadesus J, Garcia-del Portillo F. Envelope instability in DNA adenine methylase mutants of *Salmonella enterica*. Microbiology. 2002;148(4):1171–82.
Pukkila PJ, Peterson J, Herman G, Modrich P, Meselson M. Effects of high levels of DNA adenine methylation on methyl-directed mismatch repair in *Escherichia col*i. Genetics. 1983;104(4): 571–82.
Ratel D, Ravanat JL, Charles MP, Platet N, Breuillaud L, Lunardi J, Berger F, Wion D. Undetectable levels of N6-methyl adenine in mouse DNA: cloning and analysis of PRED28, a gene coding for a putative mammalian DNA adenine methyltransferase. FEBS Lett. 2006;580(13):3179–84.
Ringquist S, Smith CL. The *Escherichia coli* chromosome contains specific, unmethylated *dam* and *dcm* sites. Proc Natl Acad Sci U S A. 1992;89(10):4539–43.
Roberts D, Hoopes BC, McClure WR, Kleckner N. IS10 transposition is regulated by DNA adenine methylation. Cell. 1985;43(1):117–30.
Robertson GT, Reisenauer A, Wright R, Jensen RB, Jensen A, Shapiro L, Roop II RM. The *Brucella abortus* CcrM DNA methyltransferase is essential for viability, and its overexpression attenuates intracellular replication in murine macrophages. J Bacteriol. 2000;182(12):3482–9.
Sanchez-Romero MA, Cota I, Casadesus J. DNA methylation in bacteria: from the methyl group to the methylome. Curr Opin Microbiol. 2015;25:9–16.
Shell SS, Prestwich EG, Baek SH, Shah RR, Sassetti CM, Dedon PC, Fortune SM. DNA methylation impacts gene expression and ensures hypoxic survival of *Mycobacterium tuberculosis*. PLoS Pathog. 2013;9(7):e1003419.
Srikhanta YN, Dowideit SJ, Edwards JL, Falsetta ML, Wu HJ, Harrison OB, Fox KL, Seib KL, Maguire TL, Wang AH, Maiden MC, Grimmond SM, Apicella MA, Jennings MP. Phasevarions mediate random switching of gene expression in pathogenic *Neisseria*. PLoS Pathog. 2009;5(4):e1000400.
Srikhanta YN, Fox KL, Jennings MP. The phasevarion: phase variation of type III DNA methyltransferases controls coordinated switching in multiple genes. Nat Rev Microbiol. 2010; 8(3):196–206.
Srikhanta YN, Maguire TL, Stacey KJ, Grimmond SM, Jennings MP. The phasevarion: a genetic system controlling coordinated, random switching of expression of multiple genes. Proc Natl Acad Sci U S A. 2005;102(15):5547–51.
Stephens C, Reisenauer A, Wright R, Shapiro L. A cell cycle-regulated bacterial DNA methyltransferase is essential for viability. Proc Natl Acad Sci U S A. 1996;93(3):1210–4.
Sternberg N, Sauer B, Hoess R, Abremski K. Bacteriophage P1 *cre* gene and its regulatory region. Evidence for multiple promoters and for regulation by DNA methylation. J Mol Biol. 1986;187(2):197–212.
Tavazoie S, Church GM. Quantitative whole-genome analysis of DNA-protein interactions by in vivo methylase protection in *E. coli*. Nat Biotechnol. 1998;16(6):566–71.
Torreblanca J, Casadesus J. DNA adenine methylase mutants of *Salmonella typhimurium* and a novel dam-regulated locus. Genetics. 1996;144(1):15–26.
Urig S, Gowher H, Hermann A, Beck C, Fatemi M, Humeny A, Jeltsch A. The *Escherichia coli dam* DNA methyltransferase modifies DNA in a highly processive reaction. J Mol Biol. 2002;319(5):1085–96.
van der Woude M, Braaten B, Low D. Epigenetic phase variation of the *pap* operon in *Escherichia coli*. Trends Microbiol. 1996;4(1):5–9.
van der Woude M, Hale WB, Low DA. Formation of DNA methylation patterns: nonmethylated GATC sequences in *gut* and *pap* operons. J Bacteriol. 1998;180(22):5913–20.
van der Woude MW. Re-examining the role and random nature of phase variation. FEMS Microbiol Lett. 2006;254(2):190–7.

van der Woude MW. Phase variation: how to create and coordinate population diversity. Curr Opin Microbiol. 2011;14(2):205–11.
van der Woude MW, Bäumler AJ. Phase and antigenic variation in bacteria. Clin Microbiol Rev. 2004;17(3):581–611.
Vanyushin BF, Belozersky AN, Kokurina NA, Kadirova DX. 5-methylcytosine and 6-methylamino-purine in bacterial DNA. Nature. 1968;218(5146):1066–7.
Vanyushin BF, Buryanov YI, Belozersky AN. Distribution of N6-methyladenine in DNA of T2 phage and its host *Escherichia coli* B. Nat New Biol. 1971;230(1):25–7.
Vasu K, Nagaraja V. Diverse functions of restriction-modification systems in addition to cellular defense. Microbiol Mol Biol Rev. 2013;77(1):53–72.
Waldminghaus T, Skarstad K. The *Escherichia coli* SeqA protein. Plasmid. 2009;61(3):141–50.
Waldron DE, Owen P, Dorman CJ. Competitive interaction of the OxyR DNA-binding protein and the Dam methylase at the antigen 43 gene regulatory region in *Escherichia coli*. Mol Microbiol. 2002;44(2):509–20.
Wallecha A, Correnti J, Munster V, van der Woude M. Phase variation of Ag43 is independent of the oxidation state of OxyR. J Bacteriol. 2003;185(7):2203–9.
Wallecha A, Munster V, Correnti J, Chan T, van der Woude M. Dam- and OxyR-dependent phase variation of *agN43*: essential elements and evidence for a new role of DNA methylation. J Bacteriol. 2002;184(12):3338–47.
Wang MX, Church GM. A whole genome approach to in vivo DNA-protein interactions in *E. coli*. Nature. 1992;360(6404):606–10.
Watson Jr ME, Jarisch J, Smith AL. Inactivation of deoxyadenosine methyltransferase (dam) attenuates *Haemophilus influenzae* virulence. Mol Microbiol. 2004;53(2):651–64.
Wilson GG, Murray NE. Restriction and modification systems. Annu Rev Genet. 1991;25:585–627. doi:10.1146/annurev.ge.25.120191.003101.
Wion D, Casadesus J. N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. Nat Rev Microbiol. 2006;4(3):183–92.
Wright R, Stephens C, Shapiro L. The CcrM DNA methyltransferase is widespread in the alpha subdivision of proteobacteria, and its essential functions are conserved in *Rhizobium meliloti* and *Caulobacter crescentus*. J Bacteriol. 1997;179(18):5869–77.

Domain Structure of the Dnmt1, Dnmt3a, and Dnmt3b DNA Methyltransferases

Shoji Tajima, Isao Suetake, Kohei Takeshita, Atsushi Nakagawa, and Hironobu Kimura

Abstract

In mammals, three DNA methyltransferases, Dnmt1, Dnmt3a, and Dnmt3b, have been identified. Dnmt3a and Dnmt3b are responsible for establishing DNA methylation patterns produced through their de novo-type DNA methylation activity in implantation stage embryos and during germ cell differentiation. Dnmt3-like (Dnmt3l), which is a member of the Dnmt3 family but does not possess DNA methylation activity, was reported to be indispensable for global methylation in germ cells. Once the DNA methylation patterns are established, maintenance-type DNA methyltransferase Dnmt1 faithfully propagates them to the next generation via replication. All Dnmts possess multiple domains, and in this chapter, the structures and functions of these domains are described.

S. Tajima (✉) • H. Kimura
Laboratory of Epigenetics, Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan
e-mail: tajima@protein.osaka-u.ac.jp

I. Suetake
Laboratory of Epigenetics, Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan

CREST/AMED, 1-7-1 Otemachi, Chiyoda-ku, Tokyo 100-0004, Japan

K. Takeshita
Laboratory of Supramolecular Crystallography, Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan

PRESTO/JST, Saitama 332-0012, Japan

A. Nakagawa
Laboratory of Supramolecular Crystallography, Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan

© Springer International Publishing Switzerland 2016
A. Jeltsch, R.Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*, Advances in Experimental Medicine and Biology 945,
DOI 10.1007/978-3-319-43624-1_4

Abbreviations

AdoHcy	*S*-Adenosyl-L-homocysteine
AdoMet	*S*-Adenosyl-L-methionine
DMR	Differentially methylated region
ES cells	Embryonic stem cells
ICF syndrome	Immunodeficiency, centromeric instability, and facial anomalies syndrome
NTD	The N-terminal independently folded domain
RFTS domain	Replication foci-targeting sequence domain
SRA domain	The SET and RING-associated domain
TDG	Thymine DNA glycosylase
Tet enzyme	Ten-eleven translocation enzyme
TRD	The target recognition domain

1 DNA Methylation and Methyltransferases in Mammals

The methylation patterns of genomic DNA are established at an early stage of embryogenesis. Once the global methylation patterns are established, they are maintained during replication in a cell lineage-dependent manner (Fig. 1a). In mammals, a second methylation reprogramming occurs in gametogenesis. The global DNA methylation patterns are removed during an early stage of germ cell development and reestablished before meiosis in gonocytes in males and growing oocytes in females (Bird 2002). The expression of more than a hundred genes on autosomes is regulated in a sex-dependent manner, these genes being called imprinted genes. These genes are characterized by differentially methylated regions (DMRs), which undergo distinct DNA methylation in the male and female genomes. Generally, the DMR methylation patterns are established in germ cells at an identical stage to that of global DNA methylation (Kaneda et al. 2004). In mammals, three DNA methyltransferases, Dnmt1, Dnmt3a, and Dnmt3b, have been identified (Bestor et al 1988; Okano et al. 1998). Dnmt3a and Dnmt3b are responsible for establishing DNA methylation patterns produced through their de novo-type DNA methylation activity in implantation stage embryos and during germ cell differentiation (Okano et al 1999). Dnmt3-like (Dnmt3l), which is a member of the Dnmt3 family but does not possess DNA methylation activity, was reported to be indispensable for global methylation in germ cells (Bourc'his et al. 2001; Hata et al. 2002). Once the DNA methylation patterns are established, the maintenance-type DNA methyltransferase Dnmt1 faithfully propagates them to the next generation after DNA replication. Dnmt1 preferentially methylates hemimethylated CpG sites, which appear after DNA replication and repair.

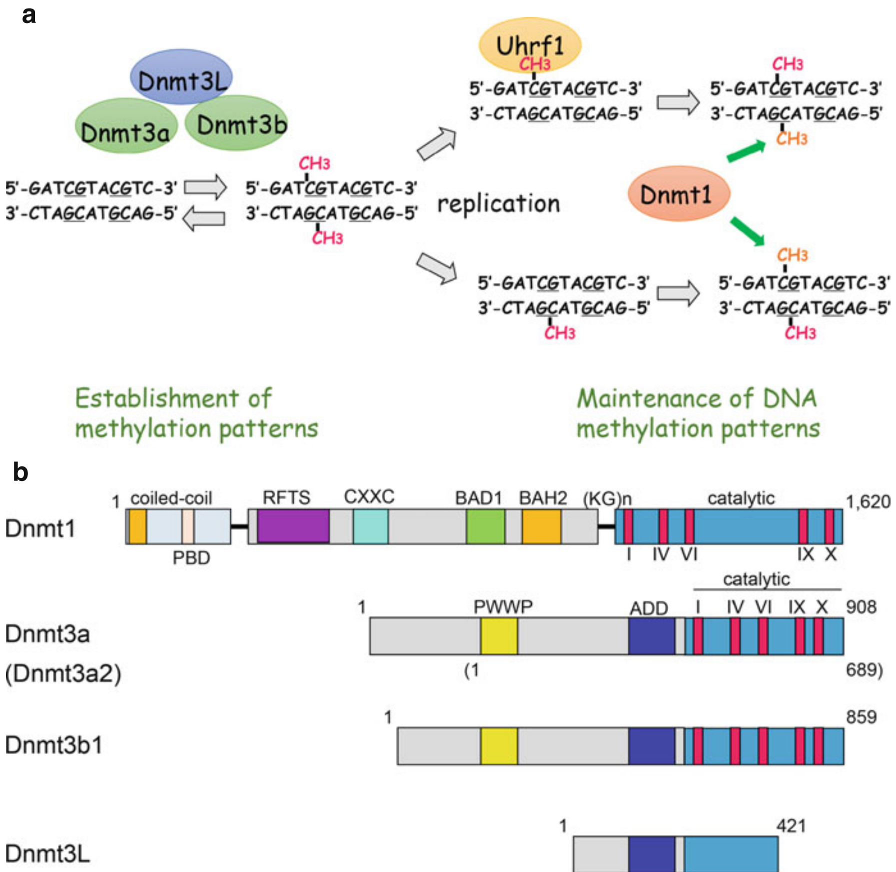


Fig. 1 Schematic illustration of establishment and maintenance of DNA methylation patterns. (**a**) The methylation patterns of genomic DNA are established at an early stage of embryogenesis by de novo-type DNA methyltransferases, Dnmt3a and Dnmt3b, with the aid of Dnmt3l. Once the global methylation patterns are established, they are maintained during replication by maintenance DNA methyltransferase Dnmt1 in collaboration with Uhrf1 in a cell lineage-dependent manner. (**b**) Schematic illustration of mammalian DNA methyltransferases, Dnmt1, Dnmt3a, and Dnmt3b. Dnmt3a has a short isoform utilizing different promoter and a transcription start site, Dnmt3a2. Dnmt3l, a member of the Dnmt3 family, lacks the catalytic domain and thus does not exhibit DNA methylation activity

2 Enzymes Responsible for the Establishment of DNA Methylation Patterns

In mammals, two of the three DNA-(cytosine C5)-methyltransferases, Dnmt3a and Dnmt3b, which are encoded in distinct gene loci, are responsible for establishing the methylation patterns through their de novo-type DNA methylation activity (Okano et al. 1999; Aoki et al. 2001). Their domain arrangements are similar, each

comprising a PWWP, ADD (Atrx-Dnmt3-Dnmt3l), and C-terminal catalytic domain (Fig. 1b). The PWWP domain is reported to bind to DNA (Qiu et al. 2002) and histone tails (Dhayalan et al. 2010) and the ADD domain to interact with various proteins including histone tails, as described below (Fuks et al. 2001; Brenner et al. 2005; Otani et al. 2009). Dnmt3l, a homologue of Dnmt3a and Dnmt3b, possesses no conserved domain for DNA methylation but contains an ADD domain (Aapola et al. 2000) and is necessary for global DNA methylation (Bourc'his et al. 2001; Hata et al. 2002).

2.1 PWWP Domain

The PWWP domain of Dnmt3 enzymes, comprising 100–150 amino acid residues, is characterized by a central core sequence motif of Pro-Trp-Trp-Pro. It was hypothesized that the domain contributes to protein-protein interactions, especially of proteins involved in cell division, growth, and differentiation, based on a comparison of 39 proteins containing a PWWP domain (Steca et al. 2000). The PWWP domain of Dnmt3b, comprising a beta-barrel structure with 5 beta-sheets followed by a five-helix bundle, turned out to be a fold responsible for DNA binding (Qiu et al. 2002). Positively charged Lys and Arg residues on the surface of the domain are expected to be the sites for DNA binding (Fig. 2a). The beta-barrel part of the PWWP domain is homologous to that of the SAND domain, which is a DNA-binding motif, and the Tudor domain, which is generally a histone-binding motif. The PWWP domain of Dnmt3a also binds to DNA, though the affinity toward DNA is one order of magnitude lower compared to that of the PWWP domain of Dnmt3b (Purdy et al. 2010).

The PWWP domains of Dnmt3a and Dnmt3b tether them to chromatin regions (Ge et al. 2004), especially to pericentric heterochromatin and thus are responsible for their DNA methylation (Chen et al. 2004). The higher affinity of Dnmt3b to DNA than that of Dnmt3a could be the reason for the specific methylation of major satellites by Dnmt3b. A point mutation in the PWWP domain in Dnmt3b was found to be the cause of the immunodeficiency, centromeric instability, and facial anomalies (ICF) syndrome (Shirohzu et al. 2002), which is the consequence of hypomethylation of the pericentromere (Okano et al. 1999; Hansen et al. 1999). Additionally, the PWWP domain interacts with tri-methylated Lys 36 histone H3 (H3K36me3), and the binding is inhibited by the point mutation causing the ICF syndrome, indicating that the recognition of H3K36me3 is crucial for Dnmt3a to target chromatin (Dhayalan et al. 2010). On the other hand, it was also reported that the PWWP domain of Dnmt3b, but not that of Dnmt3a, is recruited to the H3K36me3-containing gene body for de novo methylation (Baubec et al. 2015).

ZHX1, a member of the zinc finger and homeobox protein family, interacts with the PWWP domain of Dnmt3b and contributes to gene silencing (Kim et al. 2007). It was also reported that the PWWP domain of Dnmt3a is involved in the interaction with thymine DNA glycosylase (TDG), and the interaction apparently inhibits the DNA methylation activity of Dnmt3a (Li et al. 2007b). Since the CpG sequences

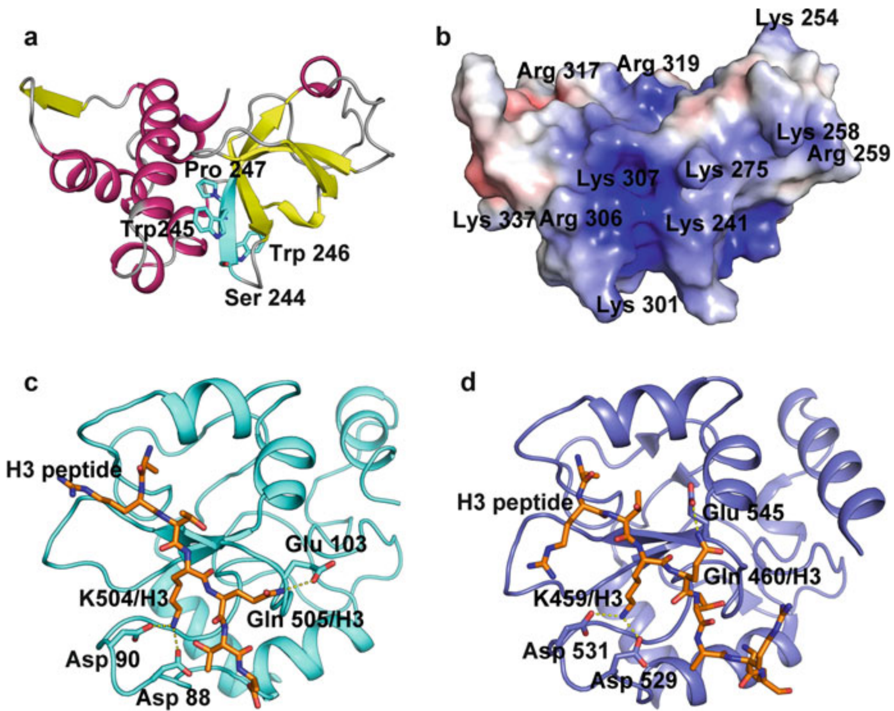


Fig. 2 Structure of the PWWP domain of Dnmt3b. (**a**) Ribbon diagram of the PWWP domain, which is changed to SWWP in Dnmt3b (*left*), and the molecular surface with the charge distribution of the PWWP domain of Dnmt3b (*right*) are shown. Positively charged amino acids (Lys and Arg) are in *blue*, negatively charged ones (Glu and Asp) in *red*, and uncharged ones in *white*. The positions of Lys and Arg residues are indicated (PDB accession number 1KHC). (**b**) Ribbon diagram of the ADD domain of Dnmt3l (PDB accession number 2PVC) (*left*) and Dnmt3a (PDB accession number 3A1B) (*right*). Histone H3 peptides are in brown. The interacting amino acid residues with histone tail peptide are conserved in Dnmt3l and Dnmt3a

methylated by Dnmt3a and Dnmt3b are the target sites for hydroxylation by an oxygenase, the ten-eleven translocation (Tet) enzyme, in embryonic stem (ES) cells (Otani et al. 2013), and TDG are proposed to contribute to the final step of the removal of oxidized methylcytosines (He et al. 2011; Maiti and Drohat 2011); the interaction between TDG and Dnmt3a indicates their strong co-relationship.

2.2 ADD Domain

The plant homeodomain (PHD)-like Atrx-Dnmt3-Dnmt3l (ADD) domain, which is rich in Cys residues, is reported to bind to many factors. The ADD domain of Dnmt3a was reported to bind to corepressor RP58 (Fuks et al. 2001), oncogene c-myc (Brenner et al. 2005), Lys 9 histone H3 (H3K9) methylase Suv39h1 and heterochromatin protein 1 (HP1) beta (Fuks et al. 2003), H3K9 methylase Setdb1 (Li et al. 2006), histone

H4 symmetrically di-methylated at Arg 3 (H4R3me2s) (Zhao et al. 2009), or histone H3 un-methylated at Lys 4 (H3K4me0) (Otani et al. 2009; Zhang et al. 2010).

The three-dimensional structure of the ADD domain of Dnmt3a is similar to those of Dnmt3l and ATRX (Argentaro et al. 2007; Ooi et al. 2007) (Fig. 2b) and possibly Dnmt3b as well (Dhayalan et al. 2011). The affinity of the ADD domain of Dnmt3a to histone H3 tail is in the sub-micromolar range and is decreased by methylation modification at Lys 4 (Otani et al. 2009). This explains why the H3K4me3, which is a mark associated with active gene promoters, protects from DNA methylation (Okitsu and Hsieh 2007; Weber et al. 2007). X-ray crystallography showed that the histone H3 tail fits into the shallow groove of the PHD finger motif in the ADD domain. The main chain of Arg 3 to Thr 6 of histone H3 forms hydrogen bonds with the ADD, and this induces a conformational change of the ADD (Otani et al. 2009). The mode of recognition of the H3K4me0 by the ADD domain of Dnmt3a is also similar to that by that of Dnmt3l, although the affinity is tenfold higher. As described below, Dnmt3l interacts directly with Dnmt3a and Dnmt3b (Suetake et al. 2004), and the proteins exist as a complex in embryonic stem (ES) cells (Li et al. 2007a). Selective recognition of H3K4me0 by the ADD domains of Dnmt3a (Dnmt3b) and Dnmt3l may recruit de novo methyltransferases to the sites to be methylated.

In plants, a DNA methyltransferase named CMT (chromomethylase) of *Arabidopsis*, which methylates the CpHpG and/or CpHpH sequence, recognizes H3K9me with its chromodomain (Stroud et al. 2014). Similar to CMT, DNA methyltransferase Dim2 of *Neurospora crassa* also contains a chromodomain and is guided to H3K9me (Tamaru and Selker 2001). For this, CMT and Dim2 cause H3K9me-dependent methylation of DNA. Although mammalian Dnmts do not directly recognize H3K9me, they are reported to interact with heterochromatin protein 1 (HP1) (Fuks et al. 2003; Smallwood et al. 2007; El Gazzar et al. 2008), which specifically recognizes H3K9me2/3. For this, H3K9 methylation is proposed to be the cause and/or result of DNA methylation.

Interestingly, the ADD domain of Dnmt3a is located at a position that inhibits accession of substrate DNA to the catalytic domain (Guo et al. 2015). The binding of the N-terminal tail of histone H3 induces rearrangement of the ADD domain to change its position to the one that DNA can access. Enhancement of de novo methylation at the chromatin region enriched in nucleosomes containing H3K4me0 reported previously (Zhang et al. 2010; Li et al. 2011) may be well explained by the conformational rearrangement of the ADD domain positioning (Guo et al. 2015) (Fig. 3). It will be important to determine whether or not the factors that are reported to interact with the ADD domain of Dnmt3a or Dnmt3b induce similar rearrangement of the enzyme to enhance de novo DNA methylation activity.

2.3 Catalytic Domain

In the catalytic domains of Dnmt3a and Dnmt3b, ten motifs characteristic for DNA-(cytosine C5)-methylation activity are conserved (Kumar et al. 1994). Dnmt3a and Dnmt3b interact through their C-terminal domains with the C-terminal domain of

Dnmt3l, and this interaction enhances de novo DNA methylation activity (Suetake et al. 2004; Chen et al. 2005). The crystal structure of the catalytic domain of the Dnmt3a in complex with the C-terminal half of Dnmt3l has been determined (Jia et al. 2007). It is a heterotetramer comprising two Dnmt3a molecules in the center and one Dnmt3l molecule at each edge (Fig. 3) (Jia et al. 2007; Jurkowska et al. 2011). The catalytic domain of Dnmt3a forms a dimer and this dimer formation is crucial for DNA methylation activity. The dimerization is expected to increase the affinity for substrate DNA as the DNA-binding site of Dnmt3a is rather small compared to that of bacterial M.HhaI. In the absence of Dnmt3l, however, Dnmt3a tends to polymerize using the same interaction surface as Dnmt3l. As the two interaction surfaces of Dnmt3a that cause polymerization contribute to its heterochromatin formation, it was proposed that formation of the complex with Dnmt3l may promote

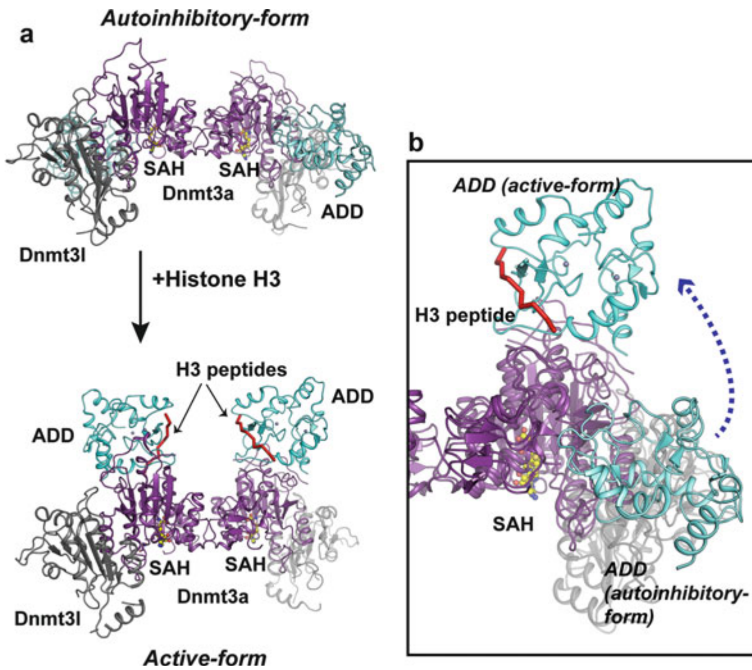


Fig. 3 Auto-inhibition of DNMT3A by the ADD domain and histone H3 tail-induced activation of DNA methylation activity. (**a**) Ribbon illustrations of the structure of the complex of the catalytic domain with the ADD domain of Dnmt3a and the C-terminal half of Dnmt3l without (*upper*) or with (*lower*) a histone H3 tail. The catalytic domain is shown in magenta, the ADD domain in cyan, and the C-terminal region of Dnmt3l in gray. *S*-Adenosyl-L-homocysteine (AdoHcy) is in yellow and the histone H3 tail in red. In the absence of a histone H3 tail, substrate DNA cannot gain access to the catalytic center as the ADD domain is in a position that inhibits the DNA binding (auto-inhibitory form; PDB accession number 4U7P). The addition of a histone H3 tail (*red*) drastically changes the position of the ADD domain to one that allows accession of DNA to the catalytic center (active form; PDB accession number 4U7T). (**b**) Superimposition of the active and auto-inhibitory forms. The *dotted arrow* indicates the movement of the ADD domain from the histone H3 tail free to the bound form

releasing Dnmt3a from heterochromatin and facilitates Dnmt3a access to the substrate DNA (Jurkowska et al. 2011). It was proposed that this inhibition of polymerization of Dnmt3a by Dnmt3l can be the underlying mechanism for the enhancement of DNA methylation activity of Dnmt3a (Jurkowska et al. 2011), especially in germ line cells to increase Dnmt3a availability and DNA methylation activity for the generation of global DNA methylation (Bourc'his et al. 2001; Hata et al. 2002; Kaneda et al. 2004). This is supported by the methylation property of Dnmt3a, which methylates periodical CpG with an 8 to 10 bp interval, this distance being estimated from the three-dimensional structure matching the DNA methylation position interval (Jia et al. 2007; Glass et al. 2009).

Interestingly, thymine DNA glycosylase (TDG), which is a T/G mismatch glycosylase, interacts with the PWWP or/and catalytic domains of Dnmt3a to modulate its DNA methylation activity (Li et al. 2006). Since methylated cytosine is converted to thymine through deamination, cytosine methylation is necessary to repair the methylation state after base excision repair. However, the physiological meaning of the interaction of Dnmt3a with TDG is rather complicated. Recently, TDG was postulated to be responsible for the removal of formylcytosine and carboxylcytosine, which are the oxidation products of methylcytosine via hydroxymethylcytosine for active demethylation initiated by Tet enzymes (He et al. 2011; Maiti and Drohat 2011).

2.4 Functions of Other Regions

An N-terminal sequence upstream of the PWWP domain in Dnmt3a, which is missing in the Dnmt3a2 isoform, strongly binds to DNA. This contributes to the DNA methylation activity and localization of the enzyme in nuclei (Suetake et al. 2011). The N-terminal region containing this DNA-binding sequence is poor in secondary structure and folding is not induced even by binding to DNA (unpublished observation). The N-terminal sequence of Dnmt3b, which exhibits no homology with that of Dnmt3a, binds to CENP-C. CENP-C is a constitutive centromere component and is necessary for mitosis. It was proposed that CENP-C recruits Dnmt3b to both centromeric and pericentromeric satellite repeats to methylate these regions (Gopalakrishnan et al. 2009). Recently, it was reported that an Arg residue in the N-terminal region undergoes citrullination by peptidylarginine deiminase 4 (PADI4), which stabilizes Dnmt3a and increases the DNA methylation level of the promoter of the *p21* gene (Deplus et al. 2014). Moreover, Dnmt3b binds to NEDD8, which is a small ubiquitin-like protein, through the region between the ADD and catalytic domains. NEDD8-modified CUL4A, which is essential for repressive chromatin formation, binds to Dnmt3b as well (Shamay et al. 2010).

2.5 Factors That Guide Dnmt3 to the Regions to Be Methylated

There have been several reports on the factors bringing Dnmt3 enzymes to specific sequences such as gene promoters. This mechanism is supported by the observation

that a short DNA sequence (methylation-determining region, MDR) can determine the DNA methylation state (Lienert et al. 2011). Sequence-specific DNA-binding proteins may recognize such a sequence. For example, Dnmt3a binds to the corepressor complex of PR48/HDAC1 or proto-oncogene c-Myc through the ADD domain (Fuks et al. 2001; Brenner et al. 2005). Dnmt3b is reported to be tethered to the centromeric and pericentromeric heterochromatin regions through interaction with CENP-C to methylate the regions (Gopalakrishnan et al. 2009). Both Dnmt3a and Dnmt3b cooperate with EVI1 (oncogene product) to bind and methylate the expression-controlling region of miRNA 124–3 (Senyuk et al. 2011). Recently, it was reported that noncoding RNA is involved in targeting of Dnmt3b to de novo methylation sites. pRNA, which binds the promoter of rRNA coding genes and forms a DNA/RNA triplex, recruits Dnmt3b to its target regions (Schmitz et al. 2010). However, it was also reported that the DNA/RNA heteroduplex rather inhibits the de novo methylation activities of both Dnmt3a and Dnmt3b in vitro (Ross et al. 2010).

In addition to the direct interaction with a DNA-binding protein or RNA, indirect interaction with the factors that bind to sequence-specific DNA-binding proteins has been reported. The KRAB zinc finger protein family, which determines target regions for methylation, comprising more than 300 genes (Liu et al. 2013), is an example. ZFP57, a KRAB zinc finger protein, binds to DNA in a sequence-specific manner and plays crucial roles in the establishment and maintenance of the methylation of imprinted genes through interaction with Trim28 (KAP1 or TIF1β) (Quenneville et al. 2011, 2012). Trim28 interacts with Dnmt3a, Dnmt3b, and Dnmt1 (Zuo et al. 2012) and acts as a scaffold to guide Dnmts to a variety of target sequences utilizing sequence-specific binding KRAB zinc finger proteins. As a similar example, NEDD8, which is an ubiquitin-like small protein modifier, acts as a tag in guiding Dnmt3b to NEDDylated proteins (Shamay et al. 2010). The main target of NEDDylation is Cullin, which plays a role in heterochromatin formation.

However, recruitment of Dnmt3a to specific genomic regions does not always introduce DNA methylation. Although Dnmt3a is recruited to a target sequence by Ezh2, a component of polycomb repressive complex 2 (PRC2) (Rush et al. 2009); MBD3, an intrinsic component of corepressor complex NuRD; Brg1, an ATPase subunit of Swi/Snf chromatin remodeling factor (Datta et al. 2005); or p53 (Wang et al. 2005), and this recruitment does not affect the DNA methylation state of the target regions.

2.6 Correlation Between de novo DNA Methylation and Histone Modifications

The PWWP domains of Dnmt3a and Dnmt3b are reported to be a motif for DNA binding (Qiu et al. 2002; Purdy et al. 2010) and bringing Dnmt3a or Dnmt3b to heterochromatin (Chen et al. 2004; Ge et al. 2004). Thus, the PWWP in the amino-terminal half of Dnmt3a or Dnmt3b is one of the determinants of methylation-site targeting. It is not known yet, however, how the PWWP of Dnmt3a or Dnmt3b selectively recognizes heterochromatin. Such recruitment of Dnmt3a or Dnmt3b to

specific regions is strongly correlated with the chromatin state or histone modifications. Trim28, which is reported to interact directly with Dnmt3a (Zuo et al. 2012), also interacts with Setdb1, a histone H3K9 methyltransferase, and HP1 (Matsui et al. 2010), which recognizes H3K9me2/3. Cullin and CENP-C, as described above, are heterochromatin finders. Many studies have shown that Dnmt3a recognizes the modified or unmodified histone tail. The PWWP domain of Dnmt3a recognizes H3K36me3 to enhance the DNA methylation activity (Dhayalan et al. 2010), and the ADD domain binds H3K4me0 (Otani et al. 2009; Li et al. 2011) to enhance the DNA methylation activity (Li et al. 2011). The histone H3 tail with K4me3 inhibits DNA methylation by Dnmt3a (Zhang et al. 2010; Li et al. 2011). Dnmt3l, a member of the Dnmt3 family with no methylation activity, also contains an ADD domain and recognizes H3K4me0 (Ooi et al. 2007). H3K4me0 recruits the Dnmt3a and Dnmt3l de novo methyltransferase complex to methylate the genome. In addition, symmetric di-methylation of Arg 3 of histone H4 (H4R3me2S) is reported to be the target of Dnmt3a via the ADD domain for DNA methylation (Zhao et al. 2009). The histone tail modifications directly recruit de novo-type Dnmt3a or Dnmt3b to the site of DNA methylation.

3 Enzymes Responsible for the Maintenance of DNA Methylation Patterns

Dnmt1 is mainly responsible for maintaining DNA methylation patterns during replication or after repair. Dnmt1 is a large molecule; mouse Dnmt1 comprises 1,620 amino acid residues. Dnmt1 is composed of several domains: the N-terminal independently folded domain (NTD), replication foci-targeting sequence (RFTS) domain, CXXC motif, two bromo adjacent homology (BAH1 and BAH2) domains, and the catalytic domain (Fig. 1b). The domains are folded almost independently and interact with each other to form a functional DNA methyltransferase. The three-dimensional structures of mouse and human Dnmt1 with all the domains except for the NTD have been reported (Takeshita et al. 2011; Zhang et al. 2015).

3.1 NTD

The NTD of mouse Dnmt1 comprising amino acids (aa) 1–248 folds independently (Suetake et al. 2006). The domain binds many factors and thus functions as a platform for the factors that regulate the Dnmt1 function. The 1–118 aa sequence in the NTD, which is a typical coiled-coil structure and is lacking in oocyte-type Dnmt1 (Mertineit et al. 1998; Gaudet et al. 1998), binds Dnmt1 associated protein 1 (DMAP1), which is a factor that represses transcription by cooperating with histone deacetylase HDAC2. DMAP1 binds to Dnmt1 at replication foci to assist maintenance of the heterochromatin state as well (Rountree et al. 2000).

Proliferating cell nuclear antigen (PCNA), which binds DNA polymerase δ and other factors related to replication, is a prerequisite factor for replication. PCNA binds to the 160–178 aa sequence of mouse Dnmt1. The binding helps Dnmt1

maintain the methylation profile of the daughter DNA (Chuang et al. 1997) and recruit Dnmt1 to replication foci at the early and middle stages of the S-phase (Schermelleh et al. 2007). Therefore, it is thought to be involved in the replication-dependent DNA methylation process. However, the NTD domain containing the PCNA-binding motif is dispensable for the maintenance of the differentially methylated regions (DMRs) of imprinted genes, at least in ES cells (Garvilles et al. 2015). The cell-cycle regulating Rb protein is also reported to bind to the NTD (Robertson et al. 2000).

Interestingly, many epigenetic factors that may contribute to the formation and maintenance of heterochromatin are reported to bind to the NTD. De novo-type DNA methyltransferases Dnmt3a and Dnmt3b (Kim et al. 2002), heterochromatin-binding protein beta (HP1 beta) that selectively recognizes H3K9me2/3 (Fuks et al. 2003), and G9a that specifically methylates H3K9 (Estève et al. 2006) bind to the NTD. All these interacting factors are related to the formation of heterochromatin, indicating that maintenance-type DNA methyltransferase Dnmt1 is tightly linked to histone methylation modification.

Although its function is not known, the NTD binds to cyclin-dependent kinase-like 5 (CDKL5) (Kameshita et al. 2008) and casein kinase (Sugiyama et al. 2010) and undergoes phosphorylation. The CDKL5 is reported to be a causative kinase for Rett syndrome. Rett syndrome is known to be caused mainly by a mutation in the *MeCP2* gene, of which the translation product specifically binds to methylated DNA and is a component of the corepressor complex. An impairment of the interaction between Dnmt1 and CDKL5 may contribute to the pathogenic process of Rett syndrome (Kameshita et al. 2008). Casein kinase 1 also interacts with the NTD. Phosphorylation with casein kinase 1 inhibits the DNA-binding activity of the NTD (Sugiyama et al. 2010). The function of the N-terminal region, which is a platform for the regulatory factors of Dnmt1, also seems to be regulated by different types of kinases (Estève et al. 2011; Lavoie and St-Pierre 2011; Lavoie et al. 2011).

In addition, the NTD of mouse Dnmt1 contains the DNA-binding 119–197 aa sequence, which overlaps with the PCNA-binding motif. The sequence contains an AT-hook-like motif and binds to the minor groove of AT-rich DNA. The DNA binding competes with the PCNA binding. Arg 133 and 136 in the sequence are crucial for the DNA-binding activity (Suetake et al. 2006). We hypothesized that this DNA-binding activity of the N-terminal domain contributes to the localization of Dnmt1 to AT-rich genome regions such as *Line1*, satellite, and the promoter of tissue-specific silent genes to maintain the fully methylated state of the repaired region that is hemimethylated (Suetake et al. 2006).

After the NTD, a flexible linker follows. Partial digestion with proteases easily releases the NTD 1–248 aa and the C-terminal part 291–1620 aa sequences (Suetake et al. 2006). According to the crystal structure of mouse Dnmt1 291–1620 aa, the structure of the RFTS domain has been determined after Pro 357 (Takeshita et al. 2011). The sequence starting from 249 to 356 aa seems to be a flexible region lacking an ordered structure. It has been reported that deletion of this region from Dnmt1 decreases maintenance methylation of the genome (Borowczyk et al. 2009). However, it has recently been reported that even with deletion of the entire NTD

including this region, Dnmt1 is fully active as a maintenance methyltransferase, at least in ES cells (Garvilles et al. 2015). The 1–353 sequence, which contains the NTD and the linker, binds to un-methylated DNA with CpG (Fatemi et al. 2001). However, the NTD contains a DNA-binding domain, which exhibits a preference not for the CpG sequence but for an AT-rich sequence (Suetake et al. 2006). The function of this linker is ambiguous at this moment.

3.2 RFTS Domain

The replication foci-targeting sequence (RFTS) domain follows the NTD. This domain is necessary for Dnmt1 localization at the replication region at the late S-phase (Leonhardt et al. 1992). This recruitment depends on the tethering of Uhrf1 (ubiquitin-like with PHD and ring finger domains 1) to the hemimethylated DNA that appears after replication, and it is a prerequisite event for the replication-dependent maintenance of DNA methylation (Bostick et al. 2007; Sharif et al. 2007). Uhrf1 selectively binds to hemimethylated DNA through the SET and RING-associated (SRA) domain (Arita et al. 2008; Avvakumov et al. 2008; Hashimoto et al. 2008), to which the RFTS domain of Dnmt1 directly binds (Berkyurek et al. 2014; Bashtrykov et al. 2014a). Direct interaction of the RFTS domain with the SRA domain accelerates the hemimethylated DNA accession to the catalytic center. The SRA of Uhrf1 and Dnmt1 cannot bind to the same CpG site at the same time due to steric hindrance (Arita et al. 2008; Song et al. 2012). This clearly indicates that there must be a mechanism to hand the hemimethylated CpG from the SRA domain over to the catalytic center of Dnmt1, which must be involved in the direct interaction between the RFTS and SRA domains (Fig. 4).

The structure of the human RFTS domain itself has been elucidated (Syeda et al. 2011) and turned out to be almost identical to that in the catalytically active mouse Dnmt1 (Takeshita et al. 2011). The position of the RFTS domain in the catalytically active Dnmt1 is intriguing. Since the RFTS domain is inserted into the catalytic pocket, a substrate DNA cannot gain access to the catalytic center due to steric hindrance. The position of the RFTS domain is stabilized by hydrogen bonds between the RFTS and catalytic domains. When the substrate DNA is short, DNA methylation activity is inhibited due to the positioning of the RFTS domain (Syeda et al. 2011; Berkyurek et al. 2014; Bashtrykov et al. 2014b). Surprisingly, even if the RFTS domain occupies the catalytic pocket, Dnmt1 can methylate DNA when it is longer than 12 bp and a length of about 30 bp is necessary for its full activity (Berkyurek et al. 2014). When the substrate DNA size is 12 bp, which is exactly the size that fits into the catalytic pocket of Dnmt1 (Song et al. 2012), Dnmt1 cannot methylate substrate DNA. The DNA methylation activity of Dnmt1 that lacks the RFTS domain toward short hemimethylated DNA is efficiently inhibited by ectopically added RFTS domain (Syeda et al. 2011; Berkyurek et al. 2014). The fact that Dnmt1 forms a head-to-head dimer through interaction between the RFTS domain

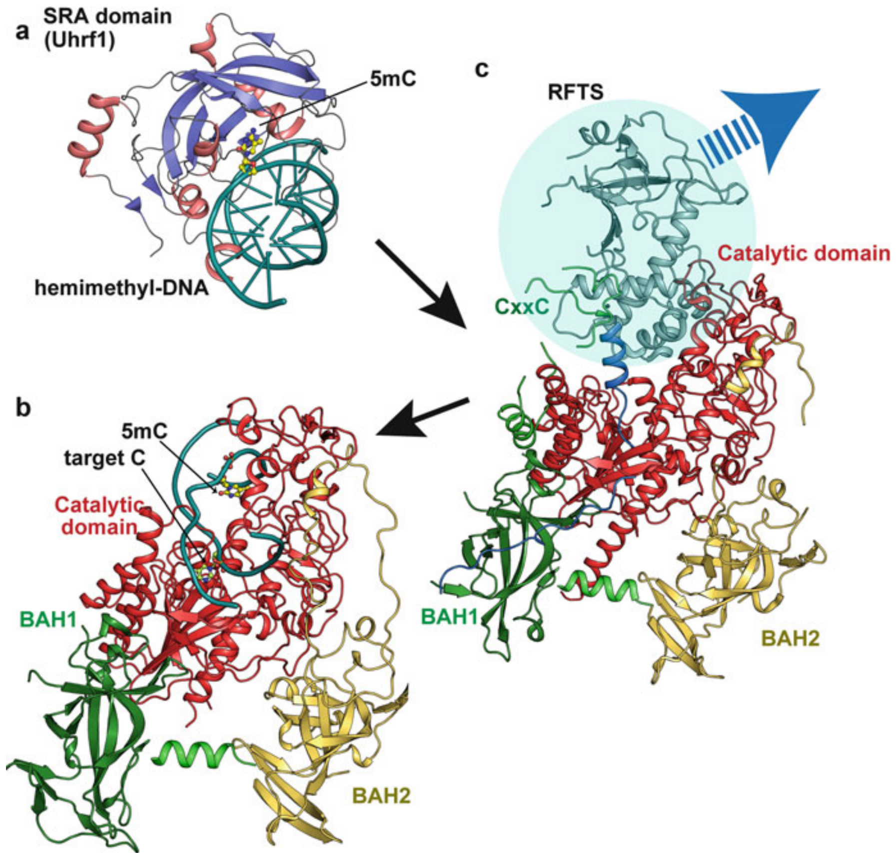


Fig. 4 Transfer of hemimethylated DNA from Uhrf1 (SRA domain) to the catalytic center of Dnmt1. (**a**) Hemimethylated DNA, which appears just after the replication, is occupied by Uhrf1 (SRA domain) (PDB accession number 2ZKE). (**b**) Considering the reported structure of the Dnmt1 catalytic domain complex with hemimethylated DNA (PDB accession number 4DA4), SRA and Dnmt1 cannot recognize hemimethylated CpG at the same time. (**c**) Furthermore, the RFTS domain occupies the catalytic pocket of Dnmt1 (PDB accession number 3AV6). The SRA domain of Uhrf1 directly interacts with the RFTS domain of Dnmt1 to remove the domain from the catalytic pocket to allow hemimethylated DNA access to the catalytic center

(Fellinger et al. 2009) may contribute to the substrate length dependence of the activity. Since the full DNA methylation activity is acquired when the substrate DNA is longer than 30 bp, the catalytic domain of Dnmt1 may increase the DNA-binding affinity by two DNA-binding sites to trigger removal of the RFTS domain from the catalytic pocket.

Amino acid residues Lys 23 (Nishiyama et al. 2013) and Lys18 of histone H3 (Qin et al. 2015) are reported to be ubiquitylated. These modifications are necessary for maintenance methylation via the interaction with the RFTS domain. Interestingly, the ring finger motif of Uhrf1, which is a prerequisite factor for

replication-dependent maintenance methylation, is involved in the ubiquitylation as an E3 ligase (Citterio et al. 2004). The tandem Tudor domain and the PHD finger of Uhrf1 recognize H3K9me3 and H3R2me0 (Arita et al. 2012), and the mutation that inhibits the recognition of H3Kme3 partly inhibits the maintenance DNA methylation (Rothbart et al. 2012) again indicating the cross talk between DNA methylation and histone modification.

Following the RFTS domain, there are three residues, Phe 631, 634, and 635, in an alpha-helix structure interacting with Tyr 1243 and Phe 1246, which are adjacent to the PCQ loop in catalytic domain motif IV, of which the Cys residue covalently binds to the target cytosine at the sixth carbon. The interactions pull the PCQ loop toward the DNA-binding pocket (Takeshita et al. 2011). Mutation of the residues decreases the DNA methylation activity (unpublished observation).

3.3 CXXC

The CXXC domain contains two zinc atoms forming zinc finger motifs, which are known to bind DNA containing un-methylated CpG. This motif is conserved among mammalian trithorax-group protein, myeloid/lymphoid leukemia (MLL) (Cierpicki et al. 2010), CXXC-type zinc finger protein 1 (CXXC1) (Voo et al. 2000), methylated DNA-binding protein 1 (MBD1) (Cross et al. 1997), and other proteins (Long et al. 2013), including Dnmt1. The CXXC domain of Dnmt1 contains two C4-type zinc fingers. The backbone structure of the CXXC domain does not change even when the RFTS domain is deleted (Takeshita et al. 2011; Hashimoto et al., PDB accession number 3SWR), or the CXXC is bound to un-methylated DNA (Song et al. 2011).

When the RFTS domain is deleted, the CXXC domain falls into the catalytic pocket, and the CXXC domain at this position binds to un-methylated DNA (Song et al. 2011). Song et al. proposed that binding of the CXXC domain to un-methylated DNA is a mechanism to inhibit its accession to the catalytic center of Dnmt1 and thus protects from methylating un-methylated DNA. If this is the case, it is reasonable to expect that deletion or mutation of the CXXC domain, which prevents the accession of un-methylated DNA to the catalytic pocket, would increase de novo-type methylation activity. However, the specificity of Dnmt1 with mutations in the CXXC domain does not cause any reduction of the genomic DNA methylation level (Frauer et al. 2011) or the specificity toward hemimethylated DNA methylation activity (Bashtrykov et al. 2012; Suetake, unpublished observation). In addition, Song et al. have reported that even Dnmt1 with the CXXC domain deleted shows similar specificity toward a hemimethylated DNA substrate (Song et al. 2012). This strongly suggests that the CXXC domain binding to un-methylated DNA does not contribute to inhibition of the methylation of un-methylated DNA. The effect of the CXXC domain mutation on the DNA methylation activity of Dnmt1 is rather controversial as Pradhan et al. reported that the CXXC is necessary for DNA methylation activity (Pradhan et al. 2008). Therefore, at present, the auto-inhibition mechanism involving the CXXC domain to prevent de novo methylation proposed by Song et al. (Song et al. 2011) needs further investigation.

The next BAH1 domain is connected with the CXXC domain by an alpha-helix structure (Takeshita et al. 2011). This helix is destroyed in the RFTS-deleted Dnmt1 without a change in the CXXC domain structure (Song et al. 2011; Takeshita et al. 2011; Hashimoto et al., PDB accession number 3SWR). A mutation or deletion of the helix changes the DNMT1 into an extended conformation and enhances the DNA methylation activity toward 12 bp DNA (Zhang et al. 2015). Since such a short DNA cannot be methylated by Dnmt1 in the absence of the SRA domain of Uhrf1 (Berkyurek et al. 2014), it is reasonable to assume that this helix region plays a crucial role in the release of the RFTS domain from the catalytic pocket.

3.4 Two BAH Domains

The CXXC domain is followed by two tandem BAH domains. The BAH domains consisting of a beta-sheet core are functionally correlated to chromatin processes. The BAH domain of RSC2, which is a component of "remodels the structure of complex" (RSC) (Chambers et al. 2013), and that of Sir3 (Armache et al. 2011; Arnaudo et al. 2013; Yang et al. 2013) interact with nucleosomes and that of ORC1, a subunit of the origin recognition complex (ORC), possesses a hydrophobic cage recognizing H4K20me2 (Kuo et al. 2012). However, the function of the two BAH domains of Dnmt1 remains unknown.

The two BAH domains of Dnmt1 are connected through an alpha-helix, which is dumbbell shaped (Takeshita et al. 2011; Song et al. 2011). At the end of the BAH1 domain, just before the helix linker, there is a zinc finger motif. The formation of this Zn-finger is necessary for a stable conformation of Dnmt1, as mutations of the involved Cys residues inhibit solubilization of the Dnmt1 protein (unpublished observation).

Interestingly, the first BAH domain (BAH1) possesses a hydrophobic cage, which is expected to recognize the methylated histone tail. Four out of six amino acid residues of the hydrophobic cage of the mouse ORC1 recognizing H4K20me2 are conserved in the BAH1 domain of Dnmt1 (Yang and Xu 2013). This may suggest that Dnmt1 interacts with the methylated histone tail in the nucleosome structure. Since the cage is masked by a long loop traversing toward the CXXC domain from the N-terminal end of the BAH1 domain, the methylation modification cannot gain access to the hydrophobic cage in this conformation (Takeshita et al. 2011). What the target of this cage structure is remains to be determined. The BAH2 domain possesses a long protruding loop from its body, of which the distal end interacts with the target recognition domain (TRD) in the catalytic domain, and adjacent residues interact directly with the substrate DNA (Song et al. 2012). Although the structure of the two BAH domains has been elucidated, their function remains elusive. Furthermore, no one has succeeded in preparing DNA methylation-active Dnmt1 without the two BAH domains suggesting that they play a crucial role in the enzyme folding or activity.

The KG-repeat between the BAH2 and catalytic domains is conserved among species (Tajima et al. 1995; Kimura et al. 1996). Until recently, this repeat was thought to be just a hinge providing flexibility to the N-terminal region and the catalytic domain. Recently, it was reported that the KG-repeat is involved in the interaction with

ubiquitin-specific protease 7 (USP7), which is a deubiquitinating enzyme (Qin et al. 2011). This interaction increases DNA methylation activity possibly through stabilizing Dnmt1 (Cheng et al. 2015). Acetylation of the Lys residues in the KG-repeat impairs the Dnmt1-USP7 interaction and promotes degradation of Dnmt1.

3.5 Catalytic Domain

Similar to other Dnmts, the ten motifs characteristic of DNA-(cytosine C5)-methyltransferases are conserved in the catalytic domain of Dnmt1. The DNA methylation mechanism of Dnmt1 is assumed to be identical to that of M.HhaI (Kumar et al. 1994). However, different from in M.HhaI (Cheng et al. 1993), the position of the side chain of Cys in the PCQ loop, which is expected to form a covalent bond with the sixth carbon of the target cytosine base (Song et al. 2012), turns toward target cytosine on the addition of methyl-group donor *S*-adenosyl-L-methionine (AdoMet) even in the absence of DNA (Takeshita et al. 2011) (Fig. 5a, c). The side chain of the Cys faces away when AdoMet is catabolized to *S*-adenosyl-L-homocysteine (AdoHcy) after the transfer of a methyl group in mouse Dnmt1 (Fig. 5b). Interestingly, the side chain of the Cys in the PCQ loop of human DNMT1 does not completely face away the side chain position even in the AdoHcy-binding form (Zhang et al. 2015) (Fig. 5d). The effect of this difference between the mouse and human enzymes remains to be determined.

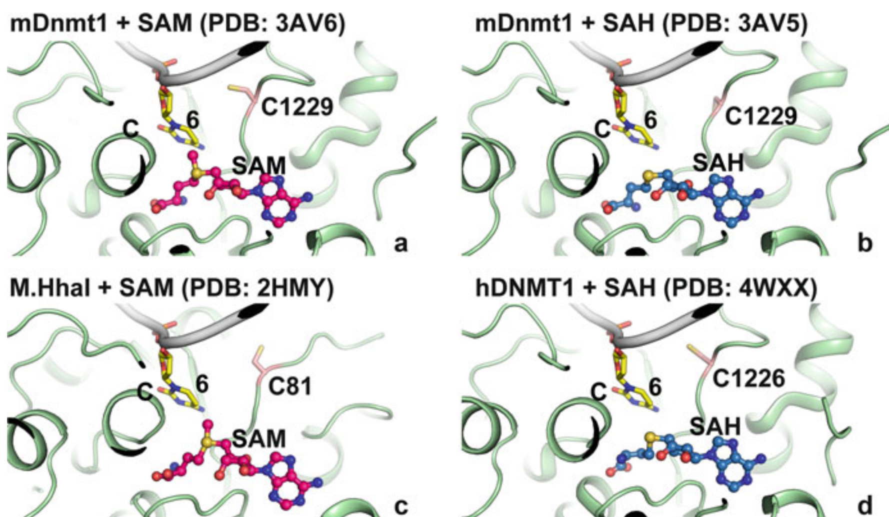


Fig. 5 Positioning of Cys residues that covalently bind the sixth position of the target cytosines in mouse Dnmt1, human DNMT1, and M.HhaI. Cys 1229 in the PCQ loop of mouse Dnmt1 faces toward or away from the target cytosine in the presence of AdoMet or AdoHcy, respectively (**a**, **b**). On the contrary, in M.Hha1, Cys 81 in the PCQ loop faces away from the target cytosine even in the presence of AdoMet when DNA is not present (**c**). Different from mouse Dnmt1, in human DNMT1, C1226 in the PCQ loop still faces toward the target cytosine in the presence of AdoHcy (**d**)

The target recognition domain (TRD) in the catalytic domain of Dnmt1 is exceptionally long compared to those in other DNA methyltransferases. The TRD covers the hemimethylated DNA and holds the methylated cytosine through hydrophobic interactions (Song et al. 2012). The target cytosine in the hemimethylated CpG is flipped out and is directly involved in the catalytic loop. According to the three-dimensional structure of the complex with hemimethylated DNA and the DNA methylation activity of the truncated Dnmt1, the recognition and selective methylation of hemimethylated DNA exist in the catalytic domain itself (Song et al. 2012; Bashtrykov et al. 2012).

4 Cross Talk Between De Novo- and Maintenance-Type DNA Methyltransferases

Establishment of DNA methylation patterns is mainly performed by de novo DNA methyltransferases, Dnmt3a and Dnmt3b, and their maintenance during replication is carried out by Dnmt1, as described above. However, it has been reported that Dnmt3a and/or Dnmt3b are also necessary for maintaining the methylation of repeat elements (Liang et al 2002). In *Dnmt3a* and *Dnmt3b* double-knockout ES cells, DNA methylation gradually decreased during culture (Chen et al. 2003). A similar decrease in DNA methylation has been observed in mouse embryonic fibroblasts after *Dnmt3b* deletion (Dodge et al. 2005). These reports indicate that not only Dnmt1 but also de novo-type DNA methyltransferases Dnmt3a and/or Dnmt3b contribute to the maintenance DNA methylation. There has been a report that Dnmt3a and Dnmt3b interact with Dnmt1 at the NTD (Kim et al. 2002). It is unlikely, however, that Dnmt3a and Dnmt3b coexist with Dnmt1 at replication foci, since Dnmt1 is loaded at an early stage of replication, and Dnmt3a and Dnmt3b at a rather late stage of replication (Alabert et al. 2014). Therefore, molecular mechanism of the cooperation with de novo-type Dnmts in maintenance DNA methylation remains to be determined.

As for the establishment of DNA methylation patterns, it was expected that Dnmt1 exhibits de novo methylation activity in vivo (Christman et al. 1995). Actually, Dnmt1 exhibits a significant level of de novo-type DNA methylation activity in vitro (Fatemi et al. 2001; Vilkaitis et al. 2005). Ectopically overexpressed Dnmt1 causes de novo DNA methylation (Takagi et al. 1995; Vertino et al. 1996; Biniszkiewicz et al. 2002). In *Dnmt3a* and *Dnmt3b* knockout ES cells, ectopically introduced DNA (Lorincz et al. 2002) as well as endogenous regions (Arand et al. 2012) undergo de novo DNA methylation. Dnmt1 apparently favors de novo methylation near preexisting methylation sites (Vilkaitis et al. 2005; Arand et al. 2012). Therefore, although the physiological meaning is elusive, Dnmt1 also causes de novo DNA methylation in vivo. The cross talk of de novo and maintenance DNA methylations is discussed in broader context in Jones and Liang (2009) and Jeltsch and Jurkowska (2014).

5 Conclusions and Perspective

Elucidation of the domain structures of Dnmts has provided important information in understanding the molecular mechanisms of DNA methylation. Indeed, the complex

of the ADD domain of Dnmt3a with histone H3 and the PWWP domain of Dnmt3b with DNA illustrated their functions in the recruitment of the enzymes to specific sites. Co-crystal structures of Dnmt3a with Dnmt3l and that of Dnmt1 with hemimethylated DNA have provided a clue to understand the DNA methylation mechanism. The domain rearrangement of Dnmt3a by histone H3 tail and occupation of the catalytic pocket of Dnmt1 by the RFTS domain have lifted the veils of DNA methylation tricks. In the near future, by utilizing the structural information, the biochemical approach with site-directed mutagenesis might provide further information in understanding molecular mechanisms of DNA methylation regulation. To this end, we need more structural information including complexes with other factors.

In addition to the high-resolution crystal structures, NMR may possibly provide us with more dynamic structural information in solution, and advanced technology of single particle analysis by electron microscopy can be a powerful technology to analyze large complexes that may be involved in DNA methylation regulation.

References

Aapola U, Kawasaki K, Scott HS, Ollila J, Vihinen M, Heino M, Shintani A, Kawasaki K, Minoshima S, Krohn K, Antonarakis SE, Shimizu N, Kudoh J, Peterson P. Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. Genomics. 2000;65:293–8.

Alabert C, Bukowski-Wills JC, Lee SB, Kustatscher G, Nakamura K, de Lima Alves F, Menard P, Mejlvang J, Rappsilber J, Groth A. Nascent chromatin capture proteomics determines chromatin dynamics during DNA replication and identifies unknown fork components. Nat Cell Biol. 2014;16:281–93.

Aoki A, Suetake I, Miyagawa J, Fujio T, Chijiwa T, Sasaki H, Tajima S. Enzymatic properties of *de novo*-type mouse DNA (cytosine-5) methyltransferases. Nucl Acids Res. 2001;29:3506–12.

Arand J, Spieler D, Karius T, Branco MR, Meilinger D, Meissner A, Jenuwein T, Xu G, Leonhardt H, Wolf V, Walter J. *In vivo* control of CpG and non-CpG DNA methylation by DNA methyltransferases. PLoS Genet. 2012;8:e1002750.

Argentaro A, Yang JC, Chapman L, Kowalczyk MS, Gibbons RJ, Higgs DR, Neuhaus D, Rhodes D. Structural consequences of disease-causing mutations in the ATRX-DNMT3-DNMT3L (ADD) domain of the chromatin-associated protein ATRX. Proc Natl Acad Sci U S A. 2007;104:11939–44.

Arita K, Ariyoshi M, Tochio H, Nakamura Y, Shirakawa M. Recognition of hemimethylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature. 2008;455:818–21.

Arita K, Isogai S, Oda T, Unoki M, Sugita K, Sekiyama N, Kuwata K, Hamamoto R, Tochio H, Sato M, Ariyoshi M, Shirakawa M. Recognition of modification status on a histone H3 tail by linked histone reader modules of the epigenetic regulator UHRF1. Proc Natl Acad Sci U S A. 2012;109:12950–5.

Armache KJ, Garlick JD, Canzio D, Narlikar GJ, Kingston RE. Structural basis of silencing: Sir3 BAH domain in complex with a nucleosome at 3.0 Å resolution. Science. 2011;334:977–82.

Arnaudo N, Fernández IS, McLaughlin SH, Peak-Chew SY, Rhodes D, Martino F. The N-terminal acetylation of Sir3 stabilizes its binding to the nucleosome core particle. Nat Struct Mol Biol. 2013;20:1119–11121.

Avvakumov GV, Walker JR, Xue S, Li Y, Duan S, Bronner C, Arrowsmith CH, Dhe-Paganon S. Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. Nature. 2008;455:822–5.

Bashtrykov P, Jankevicius G, Smarandache A, Jurkowska RZ, Ragozin S, Jeltsch A. Specificity of Dnmt1 for methylation of hemimethylated CpG sites resides in its catalytic domain. Chem Biol. 2012;19:572–8.

Bashtrykov P, Jankevicius G, Jurkowska RZ, Ragozin S, Jeltsch A. The UHRF1 protein stimulates the activity and specificity of the maintenance DNA methyltransferase DNMT1 by an allosteric mechanism. J Biol Chem. 2014a;289:4106–15.

Bashtrykov P, Rajavelu A, Hackner B, Ragozin S, Carell T, Jeltsch A. Targeted mutagenesis results in an activation of DNA methyltransferase 1 and confirms an autoinhibitory role of its RFTS domain. Chembiochem. 2014b;15:743–8.

Baubec T, Colombo DF, Wirbelauer C, Schmidt J, Burger L, Krebs AR, Akalin A, Schübeler D. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature. 2015;520:243–7.

Berkyurek AC, Suetake I, Arita K, Takeshita K, Nakagawa A, Shirakawa M, Tajima S. The DNA methyltransferase Dnmt1 directly interacts with the SET and RING finger associated (SRA) domain of the multifunctional protein Uhrf1 to facilitate accession of the catalytic center to hemi-methylated DNA. J Biol Chem. 2014;289:379–86.

Bestor T, Laudano A, Mattaliano R, Ingram V. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. J Mol Biol. 1988;203:971–83.

Biniszkiewicz D, Gribnau J, Ramsahoye B, Gaudet F, Eggan K, Humpherys D, Mastrangelo MA, Jun Z, Walter J, Jaenisch R. Dnmt1 overexpression causes genomic hypermethylation, loss of imprinting, and embryonic lethality. Mol Cell Biol. 2002;22:2124–35.

Bird A. DNA methylation patterns and epigenetic memory. Genes Dev. 2002;16:6–21.

Borowczyk E, Mohan KN, D'Aiuto L, Cirio MC, Chaillet JR. Identification of a region of the DNMT1 methyltransferase that regulates the maintenance of genomic imprints. Proc Natl Acad Sci U S A. 2009;106:20806–11.

Bostick M, Kim JK, Estève PO, Clark A, Pradhan S, Jacobsen SE. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science. 2007;317:1760–4.

Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH. Dnmt3L and the establishment of maternal genomic imprints. Science. 2001;294:2536–9.

Brenner C, Deplus R, Didelot C, Loriot A, Viré E, De Smet C, Gutierrez A, Danovi D, Bernard D, Boon T, Pelicci PG, Amati B, Kouzarides T, de Launoit Y, Di Croce L, Fuks F. Myc represses transcription through recruitment of DNA methyltransferase corepressor. EMBO J. 2005;24: 336–46.

Chambers AL, Pearl LH, Oliver AW, Downs JA. The BAH domain of Rsc2 is a histone H3 binding domain. Nucleic Acids Res. 2013;41:9168–82.

Chen T, Tsujimoto N, Li E. The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. Mol Cell Biol. 2004;24:9048–58.

Chen T, Ueda Y, Dodge JE, Wang Z, Li E. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. Mol Cell Biol. 2003;23: 5594–605.

Chen ZX, Mann JR, Hsieh CL, Riggs AD, Chédin F. Physical and functional interactions between the human DNMT3L protein and members of the de novo methyltransferase family. J Cell Biochem. 2005;95:902–17.

Cheng J, Yang H, Fang J, Ma L, Gong R, Wang P, Li Z, Xu Y. Molecular mechanism for USP7-mediated DNMT1 stabilization by acetylation. Nat Commun. 2015;6:7023.

Cheng X, Kumar S, Posfai J, Pflugrath JW, Roberts RJ. Crystal structure of the HhaI DNA methyltransferase complexed with *S*-adenosyl-L-methionine. Cell. 1993;74:299–307.

Christman JK, Sheikhnejad G, Marasco CJ, Sufrin JR. 5-Methyl-2′-deoxycytidine in single-stranded DNA can act in *cis* to signal *de novo* DNA methylation. Proc Natl Acad Sci U S A. 1995;92:7347–51.

Chuang LSH, Ian H, Koh T, Ng H, Xu G, Li BFL. Human DNA-(cytosine-5) methyltransferase PCNA complex as a target for $p21^{WAF1}$. Science. 1997;277:1996–2000.

Cierpicki T, Risner LE, Grembecka J, Lukasik SM, Popovic R, Omonkowska M, Shultis DD, Zeleznik-Le NJ, Bushweller JH. Structure of the MLL CXXC domain-DNA complex and its functional role in MLL-AF9 leukemia. Nat Struct Mol Biol. 2010;17:62–8.

Citterio E, Papait R, Nicassio F, Vecchi M, Gomiero P, Mantovani R, Di Fiore PP, Bonapace IM. Np95 is a histone-binding protein endowed with ubiquitin ligase activity. Mol Cell Biol. 2004;24:2526–35.

Cross SH, Meehan RR, Nan X, Bird A. A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. Nat Genet. 1997;16: 256–9.

Datta J, Majumder S, Bai S, Ghoshal K, Kutay H, Smith DS, Crabb JW, Jacob ST. Physical and functional interaction of DNA methyltransferase 3A with Mbd3 and Brg1 in mouse lymphosarcoma cells. Cancer Res. 2005;65:10891–900.

Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S, Jeltsch A. The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. J Biol Chem. 2010;285:26114–20.

Dhayalan A, Tamas R, Bock I, Tattermusch A, Dimitrova E, Kudithipudi S, Ragozin S, Jeltsch A. The ATRX-ADD domain binds to H3 tail peptides and reads the combined methylation state of K4 and K9. Hum Mol Genet. 2011;20:2195–203.

Deplus R, Blanchon L, Rajavelu A, Boukaba A, Defrance M, Luciani J, Rothé F, Dedeurwaerder S, Denis H, Brinkman AB, Simmer F, Müller F, Bertin B, Berdasco M, Putmans P, Calonne E, Litchfield DW, de Launoit Y, Jurkowski TP, Stunnenberg HG, Bock C, Sotiriou C, Fraga MF, Esteller M, Jeltsch A, Fuks F. Regulation of DNA methylation patterns by CK2-mediated phosphorylation of Dnmt3a. Cell Rep. 2014;8:743–53.

Dodge JE, Okano M, Dick F, Tsujimoto N, Chen T, Wang S, Ueda Y, Dyson N, Li E. Inactivation of Dnmt3b in mouse embryonic fibroblasts results in DNA hypomethylation, chromosomal instability, and spontaneous immortalization. J Biol Chem. 2005;280:17986–91.

El Gazzar M, Yoza BK, Chen X, Hu J, Hawkins GA, McCall CE. G9a and HP1 couple histone and DNA methylation to TNF alpha transcription silencing during endotoxin tolerance. J Biol Chem. 2008;283:32198–208.

Estève PO, Chang Y, Samaranayake M, Upadhyay AK, Horton JR, Feehery GR, Cheng X, Pradhan S. A methylation and phosphorylation switch between an adjacent lysine and serine determines human DNMT1 stability. Nat Struct Mol Biol. 2011;18:42–8.

Estève PO, Chin HG, Smallwood A, Feehery GR, Gangisetty O, Karpf AR, Carey MF, Pradhan S. Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. Genes Dev. 2006;20:3089–103.

Fatemi M, Hermann A, Pradhan S, Jeltsch A. The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. J Mol Biol. 2001;309:1189–99.

Fellinger K, Rothbauer U, Felle M, Längst G, Leonhardt H. Dimerization of DNA methyltransferase 1 is mediated by its regulatory domain. J Cell Biochem. 2009;106:521–8.

Fuks F, Burgers WA, Godin N, Kasai M, Kouzarides T. Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. EMBO J. 2001;20:2536–44.

Frauer C, Rottach A, Meilinger D, Bultmann S, Fellinger K, Hasenöder S, Wang M, Qin W, Söding J, Spada F, Leonhardt H. Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. PLoS One. 2011;6:e16627.

Fuks F, Hurd PJ, Deplus R, Kouzarides T. The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. Nucl Acids Res. 2003;31:2305–12.

Garvilles RG, Hasegawa T, Kimura H, Sharif J, Muto M, Koseki H, Takahashi S, Suetake I, Tajima S. Dual functions of the RFTS domain of Dnmt1 in replication-coupled DNA methylation and in protection of the genome from aberrant methylation. PLoS One. 2015;10:e0137509.

Gaudet F, Talbot D, Leonhardt H, Jaenisch R. A short DNA methyltransferase isoform restores methylation *in vivo*. J Biol Chem. 1998;273:32725–9.

Ge YZ, Pu MT, Gowher H, Wu HP, Ding JP, Jeltsch A, Xu GL. Chromatin targeting of *de novo* DNA methyltransferases by the PWWP domain. J Biol Chem. 2004;279:25447–54.

Glass JL, Fazzari MJ, Ferguson-Smith AC, Greally JM. CG dinucleotide periodicities recognized by the Dnmt3a-Dnmt3L complex are distinctive at retroelements and imprinted domains. Mamm Genome. 2009;20:633–43.

Gopalakrishnan S, Sullivan BA, Trazzi S, Della Valle G, Robertson KD. DNMT3B interacts with constitutive centromere protein CENP-C to modulate DNA methylation and the histone code at centromeric regions. Hum Mol Genet. 2009;18:3178–93.

Guo X, Wang L, Li J, Ding Z, Xiao J, Yin X, He S, Shi P, Dong L, Li G, Tian C, Wang J, Cong Y, Xu Y. Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature. 2015;517:640–4.

Hansen RS, Wijmenga C, Luo P, Stanek AM, Canfield TK, Weemaes CM, Gartler SM. The *DNMT3B* DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. Proc Natl Acad Sci U S A. 1999;96:14412–7.

Hashimoto H, Horton JR, Zhang X, Bostick M, Jacobsen SE, Cheng X. The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. Nature. 2008;455:826–9.

Hata K, Okano M, Lei H, Li E. Dnmt3L cooperates with the Dnmt3 family of *de novo* DNA methyltransferases to establish maternal imprints in mice. Development. 2002;129:1983–93.

He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, Sun Y, Li X, Dai Q, Song CX, Zhang K, He C, Xu GL. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science. 2011;333:1303–7.

Jeltsch A, Jurkowska RZ. New concepts in DNA methylation. Trends Biochem Sci. 2014;39:310–8.

Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X. Structure of Dnmt3a bound to Dnmt3L suggests a model for *de novo* DNA methylation. Nature. 2007;449:248–51.

Jones PA, Liang G. Rethinking how DNA methylation patterns are maintained. Nat Rev Genet. 2009;10:805–11.

Jurkowska RZ, Rajavelu A, Anspach N, Urbanke C, Jankevicius G, Ragozin S, Nellen W, Jeltsch A. Oligomerization and binding of the Dnmt3a DNA methyltransferase to parallel DNA molecules: heterochromatic localization and role of Dnmt3L. J Biol Chem. 2011;286:24200–7.

Kameshita I, Sekiguchi M, Hamasaki D, Sugiyama Y, Hatano N, Suetake I, Tajima S, Sueyoshi N. Cyclin-dependent kinase-like 5 binds and phosphorylates DNA methyltransferase 1. Biochem Biophys Res Commun. 2008;377:1162–7.

Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, Li E, Sasaki H. Essential role for *de novo* DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature. 2004;429:900–3.

Kim GD, Ni J, Kelesoglu N, Roberts RJ, Pradhan S. Co-operation and communication between the human maintenance and *de novo* DNA (cytosine-5) methyltransferases. EMBO J. 2002;21:4183–95.

Kim SH, Park J, Choi MC, Kim HP, Park JH, Jung Y, Lee JH, Oh DY, Im SA, Bang YJ, Kim TY. Zinc-fingers and homeoboxes 1 (ZHX1) binds DNA methyltransferase (DNMT) 3B to enhance DNMT3B-mediated transcriptional repression. Biochem Biophys Res Commun. 2007;355:318–23.

Kimura H, Ishihara G, Tajima S. Isolation and expression of a Xenopus laevis DNA methyltransferase cDNA. J Biochem. 1996;120:1182–9.

Kumar S, Cheng X, Klimasauskas S, Mi S, Posfai J, Roberts RJ, Wilson GG. The DNA (cytosine-5) methyltransferases. Nucl Acids Res. 1994;22:1–10.

Kuo AJ, Song J, Cheung P, Ishibe-Murakami S, Yamazoe S, Chen JK, Patel DJ, Gozani O. The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. Nature. 2012;484:115–9.

Lavoie G, Estève PO, Laulan NB, Pradhan S, St-Pierre Y. PKC isoforms interact with and phosphorylate DNMT1. BMC Biol. 2011;9:31.

Lavoie G, St-Pierre Y. Phosphorylation of human DNMT1: implication of cyclin-dependent kinases. Biochem Biophys Res Commun. 2011;409:187–92.

Leonhardt H, Page AW, Weier HU, Bestor TH. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell. 1992;71:865–8873.

Li BZ, Huang Z, Cui QY, Song XH, Du L, Jeltsch A, Chen P, Li G, Li E, Xu GL. Histone tails regulate DNA methylation by allosterically activating *de novo* methyltransferase. Cell Res. 2011;21:1172–81.

Li H, Rauch T, Chen ZX, Szabó PE, Riggs AD, Pfeifer GP. The histone methyltransferase SETDB1 and the DNA methyltransferase DNMT3A interact directly and localize to promoters silenced in cancer cells. J Biol Chem. 2006;281:19489–500.

Li JY, Pu MT, Hirasawa R, Li BZ, Huang YN, Zeng R, Jing NH, Chen T, Li E, Sasaki H, Xu GL. Synergistic Function of DNA Methyltransferases Dnmt3a and Dnmt3b in the Methylation of *Oct4* and *Nanog*. Mol Cell Biol. 2007a;27:8748–59.

Li YQ, Zhou PZ, Zheng XD, Walsh CP, Xu GL. Association of Dnmt3a and thymine DNA glycosylase links DNA methylation with base-excision repair. Nucl Acids Res. 2007b;35:390–400.

Liang G, Chan MF, Tomigahara Y, Tsai YC, Gonzales FA, Li E, Laird PW, Jones PA. Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. Mol Cell Biol. 2002;22:480–91.

Lienert F, Wirbelauer C, Som I, Dean A, Mohn F, Schübeler D. Identification of genetic elements that autonomously determine DNA methylation states. Nat Genet. 2011;43:1091–7.

Liu Y, Zhang X, Blumenthal RM, Cheng X. A common mode of recognition for methylated CpG. Trends Biochem Sci. 2013;38:177–83.

Long HK, Blackledge NP, Klose RJ. ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. Biochem Soc Trans. 2013;41:727–40.

Lorincz MC, Schübeler D, Hutchinson SR, Dickerson DR, Groudine M. DNA methylation density influences the stability of an epigenetic imprint and Dnmt3a/b-independent *de novo* methylation. Mol Cell Biol. 2002;22:7572–80.

Maiti A, Drohat AC. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. J Biol Chem. 2011;286:35334–8.

Matsui T, Leung D, Miyashita H, Maksakova IA, Miyachi H, Kimura H, Tachibana M, Lorincz MC, Shinkai Y. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. Nature. 2010;464:927–31.

Mertineit C, Yoder JA, Taketo T, Laird DW, Trasler JM, Bestor TH. Sex-specific exons control DNA methyltransferase in mammalian germ cells. Development. 1998;125:889–97.

Nishiyama A, Yamaguchi L, Sharif J, Johmura Y, Kawamura T, Nakanishi K, Shimamura S, Arita K, Kodama T, Ishikawa F, Koseki H, Nakanishi M. Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. Nature. 2013;502:249–53.

Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell. 1999;99:247–57.

Okano M, Xie S, Li E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet. 1998;19:219–20.

Okitsu C, Hsieh C. DNA methylation dictates histone H3K4 methylation. Mol Cell Biol. 2007;27:2746–57.

Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin SP, Allis CD, Cheng X, Bestor TH. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature. 2007;448:714–7.

Otani J, Kimura H, Sharif J, Endo TA, Mishima Y, Kawakami T, Koseki H, Shirakawa M, Suetake I, Tajima S. Cell cycle-dependent turnover of 5-hydroxymethyl cytosine in mouse embryonic stem cells. PLoS One. 2013;8:e82961.

Otani J, Nankumo T, Arita K, Inamoto S, Ariyoshi M, Shirakawa M. Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. EMBO Rep. 2009;10:1235–41.

Pradhan M, Estève PO, Chin HG, Samaranayke M, Kim GD, Pradhan S. CXXC domain of human DNMT1 is essential for enzymatic activity. Biochemistry. 2008;47:10000–9.

Purdy MM, Holz-Schietinger C, Reich NO. Identification of a second DNA binding site in human DNA methyltransferase 3A by substrate inhibition and domain deletion. Arch Biochem Biophys. 2010;498:13–22.

Qin W, Leonhardt H, Spada F. Usp7 and Uhrf1 control ubiquitination and stability of the maintenance DNA methyltransferase Dnmt1. J Cell Biochem. 2011;112:439–44.

Qin W, Wolf P, Liu N, Link S, Smets M, Mastra FL, Forné I, Pichler G, Hörl D, Fellinger K, Spada F, Bonapace IM, Imhof A, Harz H, Leonhardt H. DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination. Cell Res. 2015;25:911–29.

Qiu C, Sawada K, Zhang X, Cheng X. The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. Nat Struct Biol. 2002;9:217–24.

Quenneville S, Turelli P, Bojkowska K, Raclot C, Offner S, Kapopoulou A, Trono D. The KRAB-ZFP/KAP1 system contributes to the early embryonic establishment of site-specific DNA methylation patterns maintained during development. Cell Rep. 2012;2:766–73.

Quenneville S, Verde G, Corsinotti A, Kapopoulou A, Jakobsson J, Offner S, Baglivo I, Pedone PV, Grimaldi G, Riccio A, Trono D. In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. Mol Cell. 2011;44:361–72.

Robertson KD, Ait-Si-Ali S, Yokochi T, Wade PA, Jones PL, Wolffe AP. DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. Nat Genet. 2000;25:338–42.

Ross JP, Suetake I, Tajima S, Molloy PL. Recombinant mammalian DNA methyltransferase activity on model transcriptional gene silencing short RNA-DNA heteroduplex substrates. Biochem J. 2010;432:323–32.

Rothbart SB, Krajewski K, Nady N, Tempel W, Xue S, Badeaux AI, Barsyte-Lovejoy D, Martinez JY, Bedford MT, Fuchs SM, Arrowsmith CH, Strahl BD. Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. Nat Struct Mol Biol. 2012;19:1155–60.

Rountree MR, Bachman KE, Baylin SB. DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. Nat Genet. 2000;25:269–77.

Rush M, Appanah R, Lee S, Lam LL, Goyal P, Lorincz MC. Targeting of EZH2 to a defined genomic site is sufficient for recruitment of Dnmt3a but not *de novo* DNA methylation. Epigenetics. 2009;4:404–14.

Schermelleh L, Haemmer A, Spada F, Rösing N, Meilinger D, Rothbauer U, Cardoso MC, Leonhardt H. Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. Nucl Acids Res. 2007;35:4301–12.

Schmitz KM, Mayer C, Postepska A, Grummt I. Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. Genes Dev. 2010;24:2264–9.

Senyuk V, Premanand K, Xu P, Qian Z, Nucifora G. The oncoprotein EVI1 and the DNA methyltransferase Dnmt3 co-operate in binding and *de novo* methylation of target DNA. PLoS One. 2011;6:e20793.

Shamay M, Greenway M, Liao G, Ambinder RF, Hayward SD. *De novo* DNA methyltransferase DNMT3b interacts with NEDD8-modified proteins. J Biol Chem. 2010;285:36377–86.

Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, Endo TA, Shinga J, Mizutani-Koseki Y, Toyoda T, Okamura K, Tajima S, Mitsuya K, Okano M, Koseki H. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature. 2007;450:908–13.

Shirohzu H, Kubota T, Kumazawa A, Sado T, Chijiwa T, Inagaki K, Suetake I, Tajima S, Wakui K, Miki Y, Hayashi M, Fukushima Y, Sasaki H. Three novel DNMT3B mutations in Japanese patients with ICF syndrome. Am J Med Genet. 2002;112:31–7.

Smallwood A, Estève PO, Pradhan S, Carey M. Functional cooperation between HP1 and DNMT1 mediates gene silencing. Genes Dev. 2007;21:1169–78.

Song J, Rechkoblit O, Bestor TH, Patel DJ. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. Science. 2011;331:1036–40.

Song J, Teplova M, Ishibe-Murakami S, Patel DJ. Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. Science. 2012;335:709–12.

Steca I, Naglb SB, van Ommena GB, den Dunnena JT. The PWWP domain: a potential protein-protein interaction domain in nuclear proteins influencing differentiation? FEBS Lett. 2000;473:1–5.

Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, Patel DJ, Jacobsen SE. Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. Nat Struct Mol Biol. 2014;21:64–72.

Suetake I, Hayata D, Tajima S. The amino-terminus of mouse DNA methyltransferase 1 forms an independent domain and binds to DNA with the sequence involving PCNA binding motif. J Biochem. 2006;140:763–77.

Suetake I, Mishima Y, Kimura H, Lee YH, Goto Y, Takeshima H, Ikegami T, Tajima S. Characterization of DNA-binding activity in the N-terminal domain of the DNA methyltransferase Dnmt3a. Biochem J. 2011;437:141–8.

Suetake I, Shinozaki F, Miyagawa J, Takeshima H, Tajima S. DNMT3L stimulates DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. J Biol Chem. 2004;279:27816–23.

Sugiyama Y, Hatano N, Sueyoshi N, Suetake I, Tajima S, Kinoshita E, Kinoshita-Kikuta E, Koike T, Kameshita I. The DNA-binding activity of mouse DNA methyltransferase 1 is regulated by phosphorylation with casein kinase 1δ/ε. Biochem J. 2010;427:489–97.

Syeda F, Fagan RL, Wean M, Avvakumov GV, Walker JR, Xue S, Dhe-Paganon S, Brenner C. The replication focus targeting sequence (RFTS) domain is a DNA-competitive inhibitor of Dnmt1. J Biol Chem. 2011;286:15344–51.

Tajima S, Tsuda H, Wakabayashi N, Asano A, Mizuno S, Nishimori K. Isolation and expression of a chicken DNA methyltransferase cDNA. J Biochem. 1995;117:1050–7.

Takagi H, Tajima S, Asano A. Overexpression of DNA methyltransferase in myoblast cells accelerates myotube formation. Eur J Biochem. 1995;231:282–91.

Takeshita K, Suetake I, Yamashita E, Suga M, Narita H, Nakagawa A, Tajima S. Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). Proc Natl Acad Sci U S A. 2011;108:9055–9.

Tamaru H, Selker EU. A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. Nature. 2001;414:277–83.

Vertino PM, Yen RW, Gao J, Baylin SB. *De novo* methylation of CpG island sequences in human fibroblasts overexpressing DNA (cytosine-5-)-methyltransferase. Mol Cell Biol. 1996;16:4555–65.

Vilkaitis G, Suetake I, Klimašauskas S, Tajima S. Processive methylation of hemimethylated CpG sites by mouse Dnmt1 DNA methyltransferase. J Biol Chem. 2005;280:64–72.

Voo KS, Carlone DL, Jacobsen BM, Flodin A, Skalnik DG. Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. Mol Cell Biol. 2000;20:2108–21.

Wang YA, Kamarova Y, Shen KC, Jiang Z, Hahn MJ, Wang Y, Brooks SC. DNA methyltransferase-3a interacts with p53 and represses p53-mediated gene expression. Cancer Biol Ther. 2005;4:1138–43.

Weber M, Hellmann I, Stadler M, Ramos L, Paabo S, Rebhan M, Schübeler D. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet. 2007;39:457–66.

Yang D, Fang Q, Wang M, Ren R, Wang H, He M, Sun Y, Yang N, Xu RM. Nα-acetylated Sir3 stabilizes the conformation of a nucleosome-binding loop in the BAH domain. Nat Struct Mol Biol. 2013;20:1116–8.

Yang N, Xu RM. Structure and function of the BAH domain in chromatin biology. Crit Rev Biochem Mol Biol. 2013;48:211–21.

Zhang Y, Jurkowska R, Soeroes S, Rajavelu A, Dhayalan A, Bock I, Rathert P, Brandt O, Reinhardt R, Fischle W, Jeltsch A. Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. Nucl Acids Res. 2010;38:4246–53.

Zhang ZM, Liu S, Lin K, Luo Y, Perry JJ, Wang Y, Song J. Crystal structure of human DNA methyltransferase 1. J Mol Biol. 2015;427:2520–31.

Zhao Q, Rank G, Tan YT, Li H, Moritz RL, Simpson RJ, Cerruti L, Curtis DJ, Patel DJ, Allis CD, Cunningham JM, Jane SM. PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. Nat Struct Mol Biol. 2009;16:304–11.

Zuo X, Sheng J, Lau HT, McDonald CM, Andrade M, Cullen DE, Bell FT, Iacovino M, Kyba M, Xu G, Li X. Zinc finger protein ZFP57 requires its co-factor to recruit DNA methyltransferases and maintains DNA methylation imprint in embryonic stem cells via its transcriptional repression domain. J Biol Chem. 2012;287:2107–18.

Enzymology of Mammalian DNA Methyltransferases

Renata Z. Jurkowska and Albert Jeltsch

Abstract

DNA methylation is currently one of the hottest topics in basic and biomedical research. Despite tremendous progress in understanding the structures and biochemical properties of the mammalian DNA nucleotide methyltransferases (DNMTs), principles of their regulation in cells have only begun to be uncovered. In mammals, DNA methylation is introduced by the DNMT1, DNMT3A, and DNMT3B enzymes, which are all large multi-domain proteins. These enzymes contain a catalytic C-terminal domain with a characteristic cytosine-C5 methyltransferase fold and an N-terminal part with different domains that interacts with other proteins and chromatin and is involved in targeting and regulation of the DNMTs. The subnuclear localization of the DNMT enzymes plays an important role in their biological function: DNMT1 is localized to replicating DNA via interaction with PCNA and UHRF1. DNMT3 enzymes bind to heterochromatin via protein multimerization and are targeted to chromatin by their ADD and PWWP domains. Recently, a novel regulatory mechanism has been discovered in DNMTs, as latest structural and functional data demonstrated that the catalytic activities of all three enzymes are under tight allosteric control of their N-terminal domains having autoinhibitory functions. This mechanism provides numerous possibilities for the precise regulation of the methyltransferases via controlling the binding and release of autoinhibitory domains by protein

R.Z. Jurkowska (✉)
BioMed X Innovation Center, Im Neuenheimer Feld 583, Heidelberg D-69120, Germany
e-mail: jurkowska@bio.mx

A. Jeltsch (✉)
Institute of Biochemistry, Faculty of Chemistry, University of Stuttgart,
Pfaffenwaldring 55, Stuttgart D-70569, Germany
e-mail: albert.jeltsch@ibc.uni-stuttgart.de

© Springer International Publishing Switzerland 2016

A. Jeltsch, R.Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*,
Advances in Experimental Medicine and Biology 945,
DOI 10.1007/978-3-319-43624-1_5

factors, noncoding RNAs, or by posttranslational modifications of the DNMTs. In this chapter, we summarize key enzymatic properties of DNMTs, including their specificity and processivity, and afterward we focus on the regulation of their activity and targeting via allosteric processes, protein interactors, and posttranslational modifications.

Abbreviations

ADD domain	ATRX-DNMT3-DNMT3L domain
AdoHcy	S-Adenosyl-L-homocysteine
AdoMet	S-Adenosyl-L-methionine
BAH	Bromo-adjacent homology domain
CpG	Cytosine-guanine dinucleotide separated by a phosphate
DMAP1	DNA methyltransferase-associated protein 1
DMR	Differentially methylated region
DNMT	(Mammalian) DNA nucleotide methyltransferase
HDAC	Histone deacetylase
KG repeats	Lysine-glycine repeats
KO	Knockout
MBD	Methyl-binding domain
miRNA	MicroRNA
MTase	Methyltransferase
5mC	5-Methylcytosine
ncRNA	Non(protein)-coding RNA
PCNA	Proliferating cell nuclear antigen
PBD	PCNA-binding domain
PHD	Plant homeodomain
PTM	Posttranslational modification
RFTD	Replication foci-targeting domain
SIRT1	Sirtuin 1
SRA domain	SET- and RING-associated domain
TRD	Target recognition domain
TTD	Tandem Tudor domain
UHRF1	Ubiquitin-like with PHD and ring finger domains 1
USP7	Ubiquitin-specific peptidase 7
ZHX1	Zinc finger and homeobox protein 1

1 Introduction

The expression of genes in multicellular organisms is coordinated during development and differentiation by epigenetic information comprising DNA methylation, histone tail modifications, and noncoding RNAs (for general reviews on molecular

epigenetics, cf. Allis et al. 2015; Armstrong 2013; Bonasio et al. 2010; Jones 2012; Margueron and Reinberg 2010). In mammals, DNA methylation occurs at the C5 position of the cytosine residues, primarily in CpG dinucleotides [for general reviews on DNA methylation, cf. Jeltsch and Jurkowska 2014; Jones 2012; Jurkowska et al. 2011a; Klose and Bird 2006; Suzuki and Bird 2008, and other chapters in this book]. However, only certain CpG sites are methylated, resulting in the establishment of a tissue- and cell-type-specific pattern of DNA methylation consisting of modified and unmodified sites. In different cell types, approximately 60–80 % of all CpGs in the human genome are modified (3–8 % of all cytosines). The correct methylation pattern is essential for development, and several diseases, particularly cancer, are associated with alterations of DNA methylation [for reviews cf. Baylin and Jones 2011; Bergman and Cedar 2013; Suva et al. 2013; Hamidi et al. 2015].

In mammals, DNA methylation patterns are introduced during early development and maturation of germ cells by DNA methyltransferases DNMT3A and DNMT3B, with the help of the stimulatory factor DNMT3L. DNMT3A and DNMT3B have been traditionally designated as de novo DNA MTases, as they do not display any significant preference between hemimethylated and unmethylated DNA (Okano et al. 1998; Gowher and Jeltsch 2001). In agreement with this role, DNMT3 enzymes are highly expressed in undifferentiated cells and germ cell precursors and present at much lower levels in somatic cells. In the cell nucleus, they localize to pericentromeric heterochromatin (Chen et al. 2004; Ge et al. 2004), where they are tightly bound to nucleosomes containing methylated DNA (Jeong et al. 2009; Sharma et al. 2011).

After its establishment, the methylation pattern is perpetuated for the rest of the life of an organism, with only small tissue-specific changes. The palindromic nature of the CpG sites provides an elegant mechanism for the inheritance of the DNA methylation mark, because the methylation information is encoded in both DNA strands. During DNA replication the fully methylated CpG sites get converted to hemimethylated sites, with the parental strand carrying the original methylation mark and the daughter strand being devoid of methylation. The methylation pattern is preserved after each round of DNA replication by the maintenance methyltransferase DNMT1. This enzyme is present at the replication fork, where it works as a molecular copy machine and quickly methylates the hemimethylated CpG dinucleotides, thereby restoring the original methylation pattern [for general reviews on DNA methyltransferases, cf. Jeltsch 2002; Hermann et al. 2004a; Cheng and Blumenthal 2008; Klose and Bird 2006; Jurkowska et al. 2011a]. DNMT1 is ubiquitously and highly expressed in proliferating cells, representing the major DNA MTase activity in somatic tissues throughout mammalian development, and is present only at low levels in non-dividing cells (Robertson et al. 1999). The subnuclear localization of DNMT1 changes dynamically during cell cycle; it is diffusely distributed in the nucleus during interphase, when cells are not replicating, and localizes to the replication foci in S-phase, creating a characteristic punctuate pattern when cells actively synthesize DNA (Easwaran et al. 2004; Leonhardt et al. 1992). However, it has been recently recognized that this traditional division of tasks into de novo and maintenance methyltransferases is an oversimplification and that DNA

methylation is more correctly described as a dynamic process of ongoing methylation and demethylation and that DNMT1, DNMT3A, and DNMT3B all play roles in both de novo and maintenance methylation (Jeltsch and Jurkowska 2014). Hence, the dynamic regulation and targeting of DNMTs control the methylation state of each CpG site, thereby governing all the biological processes associated with DNA methylation. As a consequence, the complex role of DNA methylation in human biology cannot be decoded without understanding of the properties of the DNMTs, including their regulation, targeting, and interaction with other chromatin factors.

2 General Features of Mammalian DNMTs

2.1 Structure and Domain Composition

Recent structural and biochemical data on DNMTs provided compelling evidence that the arrangement of the particular domains within the methyltransferases plays a central role in the regulation of the biological functions of these enzymes. The general architecture of all three mammalian DNMTs is similar. They all are multi-domain proteins, in which two functional parts can be distinguished, a large N-terminal regulatory part and a smaller C-terminal part, required for catalysis (Fig. 1) (Jeltsch 2002; Hermann et al. 2004a; Jurkowska et al. 2011a). The N-terminal parts of variable size are entirely different between the DNMT1 and the DNMT3 family. They guide the nuclear localization of the enzymes; mediate their interaction with other proteins, with regulatory nucleic acids (like non-coding RNAs), and with chromatin modifications; are subject to posttranslational modifications; and are involved in the allosteric regulation of the enzymes' activity and specificity.

The C-terminal domains harboring the catalytic centers of the enzymes are required for binding of the S-adenosyl-L-methionine (AdoMet) cofactor and the DNA substrate. These parts contain the ten conserved amino acid motifs

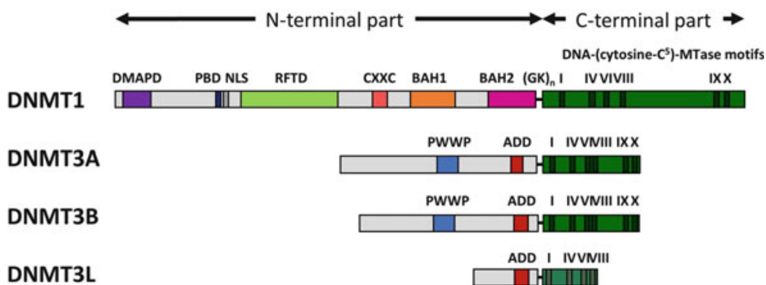


Fig. 1 Domain structure of the mammalian DNMT enzymes. Abbreviations used: *DMAPD* DNA methyltransferase-associated protein 1 interacting domain, *PBD* PCNA-binding domain, *NLS* nuclear localization signal, *RFTD* replication foci-targeting domain, *CXXC* CXXC domain, *BAH1* and *BAH2* bromo-adjacent homology domains 1 and 2, *GKn* glycine-lysine repeats, *PWWP* PWWP domain, *ADD* ATRX-DNMT3-DNMT3L domain. (With permission from Jeltsch and Jurkowska 2016, Oxford University Press)

characteristic for the common structure of all DNA-(cytosine-C5)-MTases (including bacterial enzymes and catalytic domains of eukaryotic enzymes) called the "AdoMet-dependent MTase fold," which consists of a mixed seven-stranded β-sheet, formed by six parallel β-strands and a seventh strand in an antiparallel orientation inserted into the sheet between strands five and six. Six helices surround the central β-sheet on both sides (Cheng and Blumenthal 2008; Jeltsch 2002). This domain is involved in both the cofactor binding (motifs I and X) and binding of the flipped cytosine base, followed by the methyl group transfer (motifs IV, VI, and VIII). The non-conserved region between motifs VIII and IX, the so-called target recognition domain (TRD), is involved in DNA recognition and specificity.

2.2 Catalytic Mechanism

DNA-(cytosine-C5)-methyltransferases catalyze the transfer of the methyl group from a cofactor molecule to the C5 position of cytosine residues. In this reaction, 5-methylcytosine (5mC) is created, and the AdoMet is converted to S-adenosyl-L-homocysteine (AdoHcy), which is then released from the enzyme. The transfer of the activated methyl group from AdoMet to the C5 position of the cytosine requires a close contact between the enzyme's active site and the substrate base. Such proximity is not possible while the base is located inside the DNA double helix; therefore DNA methyltransferases flip their target base out of the DNA during catalysis and bury it into a hydrophobic pocket of their active center. The base-flipping mechanism was first discovered in 1994 for the bacterial DNA C5-MTase M.HhaI (Klimasauskas et al. 1994). Later, it became clear that it is common to all DNA methyltransferases (Cheng and Roberts 2001; Jeltsch 2002). Flipping of the cytosine base was also observed in the crystal structure of DNMT1 with bound substrate DNA that has been solved recently (Song et al. 2012).

The methylation of the C5 position of cytosine is not an easy task, because cytosine is an electron-poor aromatic system and the C5 position is not intrinsically reactive, such that it will not attack the activated methylsulfonium group of AdoMet spontaneously. Therefore, a key step in the catalysis of the DNA-(cytosine-C5)-methyltransferases is the nucleophilic attack of the catalytic cysteine residue located in the PCQ motif (motif IV) on the C6 position of the cytosine ring, leading to the formation of a covalent bond between the enzyme and the substrate base. Thereby, the negative charge density at the C5 atom of the cytosine increases, so that it can attack the methyl group of the cofactor. It has been postulated that the nucleophilic attack of the cysteine might be facilitated by a transient protonation of the cytosine ring at the endocyclic nitrogen atom (N3) by an enzyme-derived acid; the conserved glutamate residue from the ENV motif (motif VI) has been proposed to carry out this reaction. In addition, the arginine residue from the RXR motif (motif VIII) may be involved in the stabilization of both the glutamate and the cytosine base. The addition of the methyl group to the base is followed by a deprotonation of the C5 atom, catalyzed by a so far unknown proton acceptor, which resolves the covalent bond between the enzyme and the base in an elimination reaction (Cheng and

Roberts 2001; Jeltsch 2002). For DNMT1, kinetic isotope effects have recently confirmed this two-step mechanism (Du et al. 2016). For DNMT3A, mutations of the key catalytic residues were found to reduce the catalytic activity, although in some cases an unexpected residual activity was observed (Reither et al. 2003; Gowher et al. 2006; Lukashevich et al. 2016).

Further mechanistic details of DNMTs, including their specificity, processivity, oligomerization, and the mechanism of DNA binding, will be discussed below for the individual enzyme families.

2.3 Regulation and Targeting of DNMTs

Despite tremendous progress in understanding the biochemical properties of the mammalian DNA methyltransferases, their regulation in cells has only begun to be uncovered. Since mammalian DNMTs do not have strong sequence specificity for target DNA regions, their targeting combined with regulation of their activity must play a central role in the generation of methylation patterns. Recently, major discoveries have been made, which showed the involvement of the N-terminal domains of the mammalian DNMTs in the enzymes' targeting and regulation. In this context, the ADD and PWWP domains of DNMT3A and DNMT3B, as well as the RFTD of DNMT1, were shown to directly bind modified histone H3 tails. Moreover, various domains (the ADD domain in DNMT3A and the CXXC and RFTD domains in DNMT1) showed an autoinhibitory interaction with the catalytic domain, demonstrating that the activity of the enzymes is under tight allosteric control. Similarly, the interaction of the N-terminal domains of DNMTs with other proteins has been shown to regulate the enzymes' activities and genome targeting. Thus, allosteric control represents a unifying concept in the regulation of DNMTs, which sets the stage for additional regulatory cues, like interacting proteins or RNAs, chromatin modifications or PTMs, which can then affect key enzymatic properties of the methyltransferases by influencing the allosteric conformation changes.

Several interaction partners of DNMTs have been described so far, for example, PCNA (Chuang et al. 1997), DNMT3L (Bourc'his et al. 2001; Hata et al. 2002; Chedin et al. 2002; Gowher et al. 2005a), UHRF1 (Sharif et al. 2007; Bostick et al. 2007), MeCP2 (Fuks et al. 2003b; Kimura and Shiota 2003) (Rajavelu et al. manuscript in preparation), or USP7 (Du et al. 2010), whose effect on the MTases has been thoroughly studied. Other important interaction partners, like HP1-beta (Fuks et al. 2003a), Mbd3 (Datta et al. 2005), the c-myc oncogene (Brenner et al. 2005); PU.1 and RP58 transcription factors (Suzuki et al. 2006; Fuks et al. 2001); the zinc finger proteins ZHX1 and Trim28 (Kim et al. 2007; Quenneville et al. 2011); the protein lysine methyltransferases (PKMTs) G9a, SUV39H1 (Fuks et al. 2003a), EZH2 (Vire et al. 2006), and SETDB1 (Li et al. 2006); and the remodeling factors HELLS (Myant and Stancheva 2008; Zhu et al. 2006), SMARCA4 (Datta et al. 2005), and hSNF2 (Geiman et al. 2004), have been reported, but their interaction with DNMTs has not yet been mechanistically investigated in detail. Finally, various aspects of the biological function of DNMTs, including their targeting and activity in cells, are regulated by posttranslational modifications (PTMs). Until now, multiple PTMs,

including phosphorylation, acetylation, ubiquitination, SUMOylation, and methylation, have been identified on mammalian DNMTs in proteomic studies. PTMs are ideally suited to mediate regulation of DNMTs' function, either by direct effects on the catalytic activity or by recruiting modification-specific readers that could then influence the enzyme's stability, activity, localization, or interaction with other proteins. Notably, the few modifications that have been functionally characterized revealed the important regulatory potential of PTMs, opening the field for future research.

Finally, non-coding RNA is an emerging player in chromatin regulation (Holoch and Moazed 2015; Rinn and Chang 2012), and RNA molecules have been shown to influence DNA methylation. In plants, a process of RNA-dependent DNA methylation exists, in which the RNA sequence directly guides DNA methylation (Matzke and Mosher 2014). Although this pathway is absent in mammals, binding of small and long non-coding RNAs to mammalian DNMTs has been shown to guide and regulate their activity. In addition, the piRNA-mediated DNA methylation in the germ line of many animals, including mammals (Iwasaki et al. 2015), recapitulates many features of an RNA-directed DNA methylation pathway, but its mechanism is not yet understood at molecular level. The direct regulation of DNA methylation by genome-encoded non-coding RNAs adds another dimension to the complex interplay between genetic information (encoded in the DNA sequence) and epigenetic information (encoded in the chromatin modification pattern, including DNA methylation).

3 Structure, Function, and Mechanism of DNMT1

3.1 Domain Composition of DNMT1

DNMT1 is a large enzyme, comprising 1620 amino acids in mouse and 1616 amino acids in human, but different isoforms of DNMT1, resulting from alternative splicing or the use of an alternative promoter, have been described (Hermann et al. 2004a; Jurkowska et al. 2011a). DNMT1 contains multiple functional domains located in the N-terminal part that is joined to the C-terminal part by a flexible linker composed of lysine-glycine (KG) repeats (Fig. 1). The N-terminal part serves as a platform for the assembly of various proteins involved in the control of chromatin structure and gene regulation.

The very N-terminus of DNMT1 contains the DMAP1 (DNA methyltransferase-associated protein 1) interaction domain that is involved in the interaction of DNMT1 with DMAP1, a transcriptional repressor, and in the stability of DNMT1 in cells. Next to it, the PCNA (proliferating cell nuclear antigen)-binding domain (PBD) has been mapped (Chuang et al. 1997). This interaction is involved in the targeting and tethering of DNMT1 to the replication fork during S-phase, which supports DNA methylation in the cell (Egger et al. 2006). The same region also contains an AT-hook like DNA-binding motif (Suetake et al. 2006). The replication foci-targeting domain (RFTD) following next is involved in the targeting of DNMT1 to replication foci (Leonhardt et al. 1992) and centromeric chromatin (Easwaran et al. 2004). This domain interacts with the UHRF1 (ubiquitin-like with PHD and

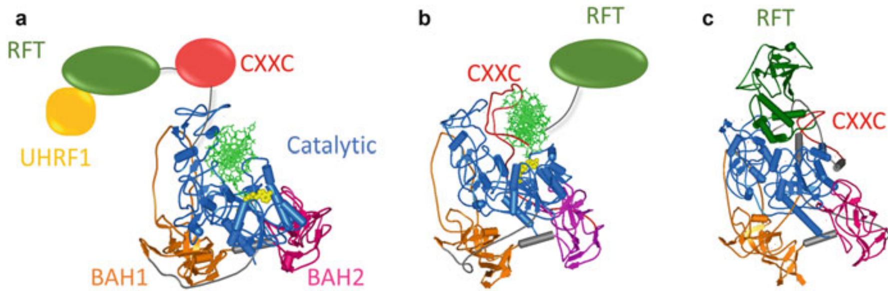


Fig. 2 Structures of DNMT1 with different N-terminal domains. (**a**) DNMT1 in active conformation with DNA (green) bound in the active site (Song et al. 2012). Removal of the autoinhibitory RFT domain can be triggered by UHRF1 (Berkyurek et al. 2014; Bashtrykov et al. 2014a). (**b**) DNMT1 with unmethylated DNA bound to the autoinhibitory CXXC domain (Song et al. 2011). (**c**) DNMT1 with the RFT domain blocking access to the active site (Takeshita et al. 2011). (With permission from Jeltsch and Jurkowska 2016, Oxford University Press)

ring finger domains 1) protein, which harbors an SRA (SET- and RING-associated) domain that specifically binds to hemimethylated DNA (see below). Moreover, it binds to ubiquitinated histone H3 tails, a modification introduced by the RING domain of UHRF1 (Nishiyama et al. 2013; Qin et al. 2015). Next, the N-terminal part of DNMT1 contains a CXXC domain that is composed of eight conserved cysteine residues that bind two zinc ions and which binds unmethylated DNA (Pradhan et al. 2008; Song et al. 2011; Bashtrykov et al. 2012a). The CXXC domain is followed by the BAH1 and BAH2 (bromo-adjacent homology 1 and 2) domains that have been proposed to act as protein-protein interaction modules; however, their functional role in DNMT1 is unknown.

The C-terminal domain of DNMT1 contains the catalytic center of the enzyme, but it is not active in an isolated form, both in vitro and in vivo, despite the presence of all motifs required for the methylation function (Fatemi et al. 2001; Margot et al. 2003). It has been shown that the N- and C- terminal parts of DNMT1 directly interact and that the catalytic domain of DNMT1 is under allosteric control of its N-terminal domain. The structural arrangement of the particular domains in DNMT1 has recently been revealed by crystallographic studies (Song et al. 2011, 2012; Takeshita et al. 2011; Syeda et al. 2011) (Fig. 2). They demonstrated that the various domains in the N-terminal part of DNMT1 surround and contact the C-terminal catalytic domain. However, all currently available structures of DNMT1 were obtained with truncated proteins, raising the possibility that they might not reflect the exact domain arrangement of the full-length protein.

3.2 Structures of DNMT1 and Allosteric Regulation

In recent years, several structures of the truncated DNMT1 proteins (lacking various parts of the N-terminus) have been solved (Song et al. 2011, 2012; Takeshita et al. 2011; Zhang et al. 2015b). They all confirmed that DNMT1 adopts the typical AdoMet-dependent MTase fold described above. These studies revealed that the

enzyme unexpectedly undergoes large domain rearrangements, which allosterically regulate its catalytic activity (Fig. 2).

A DNMT1 C-terminal fragment lacking the RFT and CXXC domains adopted an open conformation, in which the enzyme was able to bind substrate DNA and show high catalytic activity (Song et al. 2012) (Fig. 2a). This complex represented a real breakthrough in the field, as it provided the first example of a mammalian DNMT structure solved with substrate DNA bound in the active site. As expected, it showed the target base flipped out of the DNA helix and bound to DNMT1 in a manner reminiscent of other DNA MTases. In addition, this structure also showed additional unforeseen rearrangements in the DNMT1-DNA structure, including the formation of a non-Watson/Crick base pair of the orphan G residue with a G flanking the CpG site. The (then orphaned) C of the flanking G:C base pair was rotated out of the DNA helix in a direction roughly opposite to the target C flipping. However, biochemical experiments could not provide evidence for this reshuffling of base pairs and double-base flipping (Bashtrykov et al. 2012b), which suggests that the structure still may not capture the true transition state of the enzyme reaction. Moreover, the complex structure revealed several contacts of the enzyme to the target CpG site, which were validated in kinetic studies as essential for the enzyme activity and the recognition of the CpG site (described in detail in Sect. 3.3) (Bashtrykov et al. 2012b).

Furthermore, a structure of DNMT1 containing the CXXC domain showed a CpG site-specific binding of an unmethylated DNA not to the active center, but to the CXXC domain (Song et al. 2011) (Fig. 2b). This observation led to the proposal that the CXXC domain has an autoinhibitory function and it acts like a specificity filter in DNMT1 by preventing unmethylated DNA from accessing the active site. Kinetic experiments with the same DNMT1 version indeed revealed an influence of the CXXC domain on the specificity of DNMT1 (Song et al. 2011). Surprisingly, similar experiments conducted with the full-length DNMT1 did not provide evidence for a role of the CXXC domain in the specificity of DNMT1 (Bashtrykov et al. 2012a), indicating that this point deserves further attention.

Finally, another crystal structure of an almost complete DNMT1 fragment, but without DNA, provided additional seminal insight into the mechanism of DNMT1 by showing that the RFT domain binds to the active site cleft of the catalytic domain, thereby inhibiting the enzyme (Takeshita et al. 2011) (Fig. 2c). Autoinhibition was observed in biochemical studies as well (Takeshita et al. 2011; Syeda et al. 2011), and by engineering of this interface it was possible to alter the conformation of DNMT1, generating a methyltransferase that was hyperactive in vitro and in cells (Bashtrykov et al. 2014b).

Importantly, the arrangement of different domains in DNMT1 is controlled by long linker regions, which form tight interactions with surface clefts of the domains. Both the linkers and the clefts are subject to many reported PTMs in DNMT1, including phosphorylation, acetylation, and ubiquitination, which might directly control the positioning of these domains in DNMT1. Accumulating evidence indicates that the autoinhibitory mechanism of the RFT domain plays a central role as an allosteric trigger in DNMT1 (Fig. 2). Indeed, the interaction of the RFTD

with UHRF1 has been shown to stimulate the activity of DNMT1 by relieving autoinhibition (Berkyurek et al. 2014; Bashtrykov et al. 2014a), and its interaction with ubiquitinated H3 also leads to the activation of the MTase (Nishiyama et al. 2013; Qin et al. 2015).

3.3 Specificity of DNMT1

DNMT1 shows a preference for hemimethylated DNA as compared to unmethylated substrates, supporting its role as maintenance MTase (Bashtrykov et al. 2012a, b; Fatemi et al. 2001; Goyal et al. 2006; Song et al. 2012). Its intrinsic preference for hemimethylated DNA has been estimated to about 30–40-fold (Jeltsch 2006), but it depends on the exact substrate sequence, its length, and the reaction conditions. This preference has been investigated for decades, as it is one of the mechanistic foundations of the role of DNA methylation in the transfer of epigenetic information. We know now that it is molecularly based on the sequence-specific interaction of hemimethylated CpG sites with the active center of DNMT1 that is mediated by the interaction of the methyl group with a hydrophobic pocket formed by the enzyme (Song et al. 2012). More precisely, the methyl group of the 5mC is placed into a pocket formed by C1501, L1502, W1512, L1515, and M1535, which explains the preference of the enzyme for hemimethylated target sites. Further details of this process could be uncovered, if a structure of DNMT1 with an unmethylated DNA bound to the active center became available, which currently is not the case. The recognition of the 5mC-G base pair is mediated by M1535, K1537, Q1538, and R1237, which form side chain and backbone H-bonds to the edges of the base pair in the major and minor groove (Song et al. 2012). These interactions explain why the 5mC and the corresponding G in the target DNA strand are very accurately recognized by DNMT1 and why they cannot be exchanged by other nucleotides (Bashtrykov et al. 2012b).

3.4 Processivity of DNMT1

DNMT1 is a highly processive enzyme, able to methylate long stretches of hemimethylated DNA without dissociation from the substrate, a property that fits perfectly to its function as a molecular copy machine at the DNA replication fork (Goyal et al. 2006; Hermann et al. 2004b; Vilkaitis et al. 2005). Interestingly, processive methylation is restricted to one strand of the DNA, which indicates that DNMT1 does not exchange DNA strands while moving along its substrate (Hermann et al. 2004b). These biochemical findings are in a perfect agreement with the structure of DNMT1 with bound substrate DNA (Song et al. 2012), showing that the enzyme enwraps the DNA, which enables it to slide along the substrate and catalyze several successive methylation reactions without dissociation from the DNA. Due to its high processivity, DNMT1 is a very effective enzyme, ideally suited to follow DNA replication and methylate the newly synthetized DNA strand before the chromatin is reassembled.

3.5 Allosteric Regulation and Targeting of DNMT1

3.5.1 The DNMT1-PCNA Interaction

The subnuclear localization of DNMT1 changes dynamically during the cell cycle (Hermann et al. 2004a; Jurkowska et al. 2011a). It is diffusely distributed in the nucleus during interphase when cells are not replicating. In the early and mid-S-phase, DNMT1 localizes to the replication foci in cells actively synthesizing DNA, creating a characteristic punctuate pattern. Three regions of DNMT1 have been implicated in the targeting of the enzyme to the replication foci during S-phase, namely, the PCNA-binding domain (PBD) (Chuang et al. 1997), the replication foci-targeting domain (RFTD) (Leonhardt et al. 1992), and the BAH domains (Liu et al. 1998) (Fig. 2). However, deletion of the RFTD or BAH domains did not affect the delivery of DNMT1 to the replication fork (Easwaran et al. 2004), suggesting that the PBD has a central role in this process. Through this domain, DNMT1 directly interacts with PCNA, the so-called processivity factor of the replication machinery that forms a ring around the DNA helix. In addition, both proteins co-localize in vivo, indicating that PCNA might recruit DNMT1 to the replication fork and load it onto DNA. Indeed, the expression of a truncated DNMT1, lacking parts of the PBD, led to a delay in the re-methylation of DNA after replication (Egger et al. 2006). However, it did not cause massive defects in DNA methylation, indicating that the interaction of PCNA with DNMT1 contributes to the efficiency of DNA re-methylation, but it is not absolutely necessary for this process. In addition, in vitro experiments provided evidence that the interaction with PCNA increases DNA binding and catalytic activity of DNMT1 (Iida et al. 2002), uncovering the first molecular details of this process.

During progression of the S-phase, the subnuclear pattern of DNMT1 localization changes from small, punctuate, and abundant structures in early S-phase to fewer, large, toroidal structures in late S-phase, which co-localized with late replicating heterochromatic satellite DNA (Leonhardt et al. 1992; O'Keefe et al. 1992; Easwaran et al. 2004). In addition, some DNMT1 remains associated with centromeric heterochromatin in the late S- and G2-phases, even after heterochromatin replication. The interaction with heterochromatin is mediated by the PBD and occurs in a replication-independent manner (Easwaran et al. 2004), and it is at least in part mediated by the UHRF1 protein, as described in the next paragraph.

3.5.2 The DNMT1-UHRF1 Interaction

Another pathway of DNMT1 targeting has been discovered more recently with the finding that UHRF1 is required for maintaining DNA methylation in mammals (Bostick et al. 2007; Sharif et al. 2007). UHRF1 specifically binds to hemimethylated DNA via its SET- and RING-associated (SRA) domain (Bostick et al. 2007; Avvakumov et al. 2008; Hashimoto et al. 2008; Arita et al. 2008), and its localization to replicating heterochromatin is dependent on the presence of hemimethylated DNA (Sharif et al. 2007). UHRF1 co-localizes with DNMT1 and PCNA at replicating heterochromatic regions during mid- to late S-phase, and DNMT1 association with chromatin is lost in UHRF1 knockout (KO) cells (Sharif et al. 2007; Bostick et al. 2007). Remarkably, the phenotype of the UHRF1 KO in mice mimics that of the

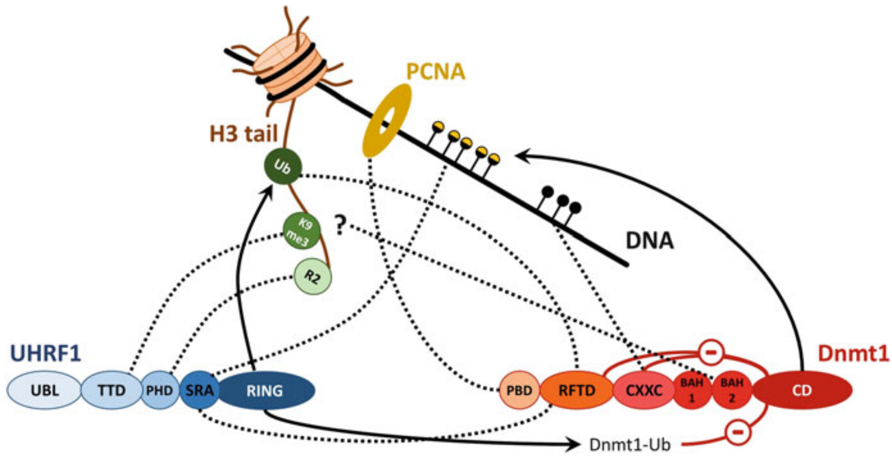


Fig. 3 Regulatory mechanisms controlling the activity and stability of DNMT1. The figure illustrates the complex interplay between DNMT1, UHRF1, and chromatin. Enzymatic activities are indicated by *arrows*. Binding ("reading") interactions are symbolized by *dotted lines*. For details cf the text

DNMT1 KO, as UHRF1-deficient embryos die shortly after gastrulation and show significantly reduced levels of DNA methylation (Sharif et al. 2007), indicating that UHRF1 has a central role in the maintenance of DNA methylation. These data led to a model that UHRF1 recruits DNMT1 to replicated hemimethylated DNA to facilitate its efficient re-methylation (Jeltsch 2008) (Fig. 3).

Later, it was found that two reading domains of UHRF1 recognize histone marks: the tandem Tudor domain (TTD) of UHRF1 binds methylated lysine 9 and unmethylated lysine 4 on histone 3 tail (Nady et al. 2011; Rothbart et al. 2012) and the plant homeodomain (PHD) of UHRF1 binds to unmodified arginine 2 of the H3 tail (Hu et al. 2011; Rajakumara et al. 2011; Wang et al. 2011). The interaction with H3K9me3 is required for the localization of UHRF1 to heterochromatin and for maintenance DNA methylation, since a mutation in TTD, which prevents binding to H3K9me3, abolished both functions (Nady et al. 2011; Rothbart et al. 2012). Similarly, disruption of H3R2 binding abolished DNA methylation by DNMT1 in cells (Qin et al. 2015). These data indicate that the coordinated recognition of two histone marks, H3K9me3 and H3R2, and the interaction with hemimethylated DNA by UHRF1 are all necessary for the guidance of DNMT1 and faithful maintenance of DNA methylation (Rothbart et al. 2013; Liu et al. 2013) (Fig. 3).

In addition to its role in targeting of DNMT1, UHRF1 was also shown to stimulate the catalytic activity of DNMT1 directly, by interacting with the RFT domain of DNMT1 and preventing the autoinhibitory conformation (Berkyurek et al. 2014; Bashtrykov et al. 2014a). Moreover, the RING domain of UHRF1 was shown to ubiquitinate H3 at K18 and K23 (Nishiyama et al. 2013; Qin et al. 2015). Ubiquitinated H3 is bound by DNMT1, leading to the stimulation of its methyltransferase activity. In addition, UHRF1 is involved in the ubiquitination of DNMT1,

which reduces DNMT1′s stability (see below). All these observations demonstrate that UHRF1 is a multifaceted regulator of DNMT1 and the entire maintenance DNA methylation machinery (Fig. 3).

3.5.3 Binding of the DNMT1-RFTD to H3 Ubiquitination

Recent studies demonstrated that UHRF1-dependent ubiquitination of histone H3 has an essential role in DNMT1 function, as a catalytically inactive RING domain mutant of UHRF1 failed to recruit DNMT1 to the replication sites (Nishiyama et al. 2013). The molecular mechanism of this observation has begun to be uncovered with the finding that DNMT1 preferentially associates with ubiquitinated H3 though its replication foci-targeting domain and that this interaction leads to the activation of the enzyme (Nishiyama et al. 2013; Qin et al. 2015). The ubiquitination of the H3 tail is introduced by the RING domain of UHRF1, which is an E3 ligase (Nishiyama et al. 2013; Qin et al. 2015). Ubiquitination has been observed at K18 and K23 of H3 and also on DNMT1 itself, but it is unclear at present how these findings are related. These data indicate an important additional connection between the DNMT1 and UHRF1 chromatin interactions, which is essential for an efficient maintenance methylation to occur (Fig. 3).

3.5.4 Regulation of Activity and Specificity of DNMT1 by Nucleic Acid Binding

DNMT1 has been shown to possess multiple DNA-binding sites, which contribute to the allosteric regulation of its activity and specificity. Many groups have reported that the enzyme shows reduced specificity in the presence of methylated DNA (Fatemi et al. 2001, 2002; Christman et al. 1995; Bacolla et al. 1999). This effect was due to an increase in the rate of de novo methylation of unmodified DNA, while the methylation of hemimethylated DNA was weakly inhibited (Fatemi et al. 2001; Goyal et al. 2006). The increase in the methylation efficiency of unmethylated DNA indicates that the binding of the methylated DNA to the N-terminal domain of the enzyme induces an allosteric activation for the methylation of unmethylated substrates. The molecular mechanism of the allosteric activation of DNMT1 is unknown; the CXXC domain (Fatemi et al. 2001) and the residues 284–287 of the murine DNMT1 (Pradhan and Esteve 2003) have been implicated in this process. In addition, an inhibitory effect of unmethylated substrates was demonstrated in several studies (Svedruzic and Reich 2005; Flynn et al. 2003; Bacolla et al. 1999), suggesting that binding of an unmethylated DNA to the N-terminal part of DNMT1 leads to a repression of the enzymatic activity on hemimethylated DNA. The binding site for this substrate inhibition effect was localized in the first 501 amino acids of DNMT1 (Bacolla et al. 2001). Additional evidence suggests that binding of the methylated DNA to the N-terminal inhibition site also caused derepression of the enzyme (Bacolla et al. 2001). Whether the inhibition and stimulation effects observed in these various studies are due to binding to the same or different sites and to what extent different DNAs compete for the different sites is not clear at present.

Interestingly, all studies agree that binding to unmethylated DNA at a secondary site reduces the activity of DNMT1, while binding to methylated DNA increases its activity. This observation could be related to the fact that DNA methylation patterns in the human genome are highly bimodal (Eckhardt et al. 2006; Meissner et al. 2008; Zhang et al. 2009), meaning that the genomic regions tend to be either highly methylated or completely unmethylated. The occurrence of the bimodal methylation patterns could be explained by the allosteric binding of the substrate DNA to a secondary site, because DNMT1 would be activated on methylated regions and inactivated on unmethylated DNA. Consequently, highly methylated regions will tend to gain methylation, whereas low methylated regions will tend to lose even their residual methylation.

In addition to DNA, DNMT1 has been shown to bind different RNAs. Initial studies showed that DNMT1 purified from insect cells contains inhibitory RNA (Glickman et al. 1997a). Later, it had been discovered that RNA binding regulates the activity of DNMT1 in a locus-specific manner. A long non-coding RNA (ncRNA) originating from the CEBPA locus was observed to bind and inhibit DNMT1 and prevent the methylation of this locus. Similar effects were observed for several other loci on a genomic scale (Di Ruscio et al. 2013). Based on these findings, the authors proposed a model, in which the ncRNAs transcribed at one locus function as a shield for this locus preventing its methylation. Thereby, the expression of the locus would be perpetuated. Recently, it was also reported that DNMT1 binds to miRNAs like miR-155-5p. Similarly as the long non-coding RNAs, miRNAs function as inhibitors of DNMT1, and the transfection of miRNAs to cells caused changes in the cellular methylation (Zhang et al. 2015a). RNA binding was mapped to the catalytic domain of DNMT1 (Di Ruscio et al. 2013; Zhang et al. 2015a), and it was reported that miRNAs act as DNA-competitive inhibitors (Zhang et al. 2015a). These findings suggest that the inhibition of DNMT1 by RNAs is based on a direct competition of the RNA inhibitor and DNA substrate for access to the catalytic center.

However, it is well conceivable that the additional DNA-binding sites described above bind regulatory RNAs as well. These important features of the interaction of DNMT1 with regulatory DNA and RNA are not yet understood at a molecular level and deserve additional experimental work.

3.6 PTMs of DNMT1

3.6.1 Phosphorylation of DNMT1

DNMT1 is subject to several posttranslational modifications like phosphorylation, methylation, ubiquitination, acetylation, and SUMOylation. Following the initial identification of S515 as a major phosphorylation site in DNMT1 purified from insect cells (Glickman et al. 1997b), several more phosphorylated serine and threonine residues have been identified in targeted and high-throughput proteomic approaches with DNMT1 purified from human cells. Currently, >60 phosphorylation sites have been mapped on human and mouse DNMT1 (http://www.phosphosite.org), but only few of them have been functionally studied. The phosphorylated

S515 is involved in the interaction between the N-terminal and catalytic domains of DNMT1 that is necessary for the activity of the enzyme (Goyal et al. 2007). Phosphorylation of S146 introduced by casein kinase 1 delta/epsilon decreases the DNA-binding affinity of DNMT1 (Sugiyama et al. 2010), and phosphorylation of S127 and S143 regulates the interaction of DNMT1 with PCNA and UHRF1 (Hervouet et al. 2010). Moreover, phosphorylation of DNMT1 by PKC has been reported, but the target sites have not been identified yet (Lavoie et al. 2011). The S143 of DNMT1 is phosphorylated by AKT1, which leads to the stabilization of the methyltransferase (Esteve et al. 2011). Recently, a specific 14-3-3 family reader protein for this modification has been identified (Esteve et al. 2016). It binds to the phosphorylated DNMT1, leading to the inhibition of DNMT1 activity, aberrant DNA methylation, and cell invasion (Esteve et al. 2016). The functional significance of many of the other phosphorylations in DNMT1 still a waits to be elucidated. In particular, the influence of the PTMs on the allosteric regulation of DNMT1 activity and specificity needs to be studied.

3.6.2 Acetylation of DNMT1

Multiple acetylation sites have been identified on DNMT1 up to date in proteomic analyses (Kim et al. 2006; Choudhary et al. 2009; Peng et al. 2011) (http://www.phosphosite.org); however, their functional significance has only begun to be revealed. Initial experiments with deacetylase inhibitors demonstrated the involvement of acetylation in the control of DNMT1 stability (Zhou et al. 2008; Peng et al. 2011). Recently, an elegant mechanism regulating the abundance of DNMT1 during the cell cycle has been identified. It starts with the acetylation of DNMT1 by the acetyltransferase Tip60, in the KG linker, followed by UHRF1-mediated ubiquitination, resulting in proteasomal degradation of DNMT1 at the end of DNA replication. In turn, histone deacetylase 1 (HDAC1) and deubiquitinase ubiquitin-specific peptidase 7 (USP7, also known as HAUSP) have an opposite effect and increase the stability of DNMT1 (Du et al. 2010; Qin et al. 2011). The crystal structure of DNMT1 in complex with USP7 has been solved and revealed that this interaction is dependent on the KG linker of DNMT1, explaining why acetylation of this region impairs complex formation and promotes degradation of DNMT1 (Cheng et al. 2015). In addition, it has been shown that SIRT1 deacetylates DNMT1 at several sites and thereby regulates the activity and function of the methyltransferase (Peng et al. 2011).

3.6.3 Lysine Methylation of DNMT1

DNMT1 is methylated by SET7/9, both in vivo and in vitro. The monomethylation of human DNMT1 by SET7/9 occurs at K142 mainly during late S-phase and it promotes proteasomal degradation of the enzyme in a cell cycle-dependent manner (Esteve et al. 2009). The methylation level is higher in the absence of the LSD1 lysine demethylase, suggesting that the K142 methylation of DNMT1 is reversible and can be removed by LSD1 (Wang et al. 2009). In addition, it is antagonistic with the phosphorylation of DNMT1 at S143 by AKT1 kinase described above (Esteve et al. 2011).

4 Structure, Function, and Mechanism of DNMT3 Enzymes

4.1 Domain Composition of DNMT3 Proteins

The DNMT3 family contains three members: DNMT3A, DNMT3B, and DNMT3L, which in human comprise 912 aa, 853 aa, and 387 aa, respectively (Fig. 1). Several isoforms of DNMT3A and DNMT3B, resulting from alternative splicing or the use of alternative start codons, have been identified both in mice and human (Jurkowska et al. 2011a). Besides the C-terminal domain required for catalysis, DNMT3A and DNMT3B possess an N-terminal domain involved in the targeting of the enzymes to chromatin and regulation of their function (Jurkowska et al. 2011a). In this part, two functional domains are present: the ADD (ATRX-DNMT3-DNMT3L) domain, also known as PHD (plant homeodomain), and the PWWP domain, which is missing in DNMT3L.

The ADD domain is a cysteine-rich region that binds zinc ions and creates a platform for protein-protein interactions. This domain mediates the interaction of DNMT3 enzymes with histone H3 tails unmethylated at lysine K4 (Ooi et al. 2007; Otani et al. 2009; Zhang et al. 2010; Guo et al. 2015). In addition, it is involved in the interaction of DNMT3A with various components of the epigenetic machinery, like protein lysine methyltransferases SUV39H1 (Fuks et al. 2003a), SETDB1 (Li et al. 2006), E2H2 (Vire et al. 2006), and deacetylase HDAC1; reading domain proteins, including HP1ß (Fuks et al. 2003a), Mbd3 (Datta et al. 2005), and MeCP2 (Kimura and Shiota 2003; Fuks et al. 2003b; Rajavelu et al., manuscript in preparation), as well as transcriptional factors PU.1 (Suzuki et al. 2006), Myc (Brenner et al. 2005), and RP58 (Fuks et al. 2001); and remodeling factors hSNF2 (Geiman et al. 2004) and SMARCA4 (Datta et al. 2005). The ADD domain has been implicated in the allosteric control of DNMT3A, as it interacts with the catalytic domain of the methyltransferase and inhibits its activity (see below), indicating that ADD-mediated interactions with other proteins could have direct regulatory effects on the catalytic activity of the MTase.

The PWWP domain of DNMT3A and DNMT3B is a conserved region of 100–150 amino acids, containing a conserved proline-tryptophan motif (hence the name PWWP). PWWP domains belong to the royal domain superfamily, members of which interact with histone tails in various modification states (Qin and Min 2014). It has been shown that the PWWP domains of DNMT3A and DNMT3B specifically recognize the H3K36 trimethylation mark (Dhayalan et al. 2010; Baubec et al. 2015). These domains are essential for the targeting of DNMT3 enzymes to pericentromeric chromatin (Chen et al. 2004; Ge et al. 2004). The structures of the PWWP domains from both DNMT3A and DNMT3B have been solved (Qiu et al. 2002; Rondelet et al. 2016). It has been suggested that the PWWP domain would synergistically bind both H3K36me3 histone tail and DNA through its conserved aromatic cage and a positively charged surface, respectively (Qin and Min 2014; Qiu et al. 2002; Rondelet et al. 2016). In addition, the ZHX1 (zinc finger and homeobox protein 1) interacts with the PWWP domain of DNMT3B and enhances DNMT3B-mediated transcriptional repression (Kim et al. 2007).

The part of DNMT3A and DNMT3B N-terminal to the PWWP domain is the least conserved region between both enzymes. Consequently, this part may be responsible for targeting of the enzymes to different genomic loci. This domain was shown to bind DNA (Suetake et al. 2011), and it is important for anchoring the enzymes to nucleosomes (Jeong et al. 2009; Baubec et al. 2015). However, up to date no molecular or biological function has been assigned to this part.

The C-terminal domains of DNMT3A and DNMT3B, which enclose the catalytic centers of the enzymes, share approximately 85 % sequence homology, and, in contrast to the catalytic domain of DNMT1, they are active in an isolated form (Gowher and Jeltsch 2002) and have been used as a model system to study the catalytic mechanism and specificity of the DNMT3 proteins. Interestingly, isolated catalytic domains of DNMT3A and DNMT3B show higher enzymatic activity than the full-length proteins, indicating that the N-terminal domains allosterically inhibit the activity of the enzymes (Li et al. 2011). It has been shown recently that the ADD domain of DNMT3A, which directly interacts with the catalytic domain of the methyltransferase, is responsible for this inhibition in the absence of histones (Guo et al. 2015) (Fig. 4). This model is further supported by kinetic experiments, showing that binding of ADD domain of DNMT3A to H3 tail stimulates the activity of the enzyme (Guo et al. 2015; Li et al. 2011; Zhang et al. 2010).

DNMT3L, the third member of the DNMT3 family, lacks parts of the N-terminal region including the PWWP domain. Strikingly, it carries amino acid exchanges and deletions within the conserved DNA-(cytosine-C5)-MTase motifs, which include the catalytic residues, indicating that while it still adopts the typical AdoMet-dependent MTase fold described above, it cannot have catalytic activity and is unable to bind AdoMet.

4.2 Structures of DNMT3A and Allosteric Regulation

Up to date, structures containing truncated DNMT3A/DNMT3L complexes (Jia et al. 2007; Guo et al. 2015) and one structure of DNMT3L (Ooi et al. 2007) have been solved. In addition, the structures of isolated ADD and PWWP domains in free and peptide-bound forms have been obtained as well (Otani et al. 2009; Qiu et al. 2002; Rondelet et al. 2016). The structure of the complex of the C-terminal domains of DNMT3A/DNMT3L has been solved in 2007, being the first structure published for a mammalian DNMT. It showed that the complex forms a linear heterotetramer consisting of two DNMT3L (at the edges of the tetramer) and two DNMT3A molecules (in the center) (Jia et al. 2007) (Fig. 4). The heterotetrameric structure of the complex was confirmed in solution (Jurkowska et al. 2008). The structure also revealed that the C-terminal domain of DNMT3A contains two interfaces for protein-protein contacts: a hydrophobic one generated by the stacking interaction of two phenylalanine residues (called FF interface), which mediates the DNMT3A/DNMT3L interaction, and a polar interface generated by a hydrogen-bonding network between arginine and aspartate residues from both subunits (called RD interface), which can only mediate DNMT3A/DNMT3A interactions, since the corresponding region is absent in DNMT3L. DNA-binding studies showed that the

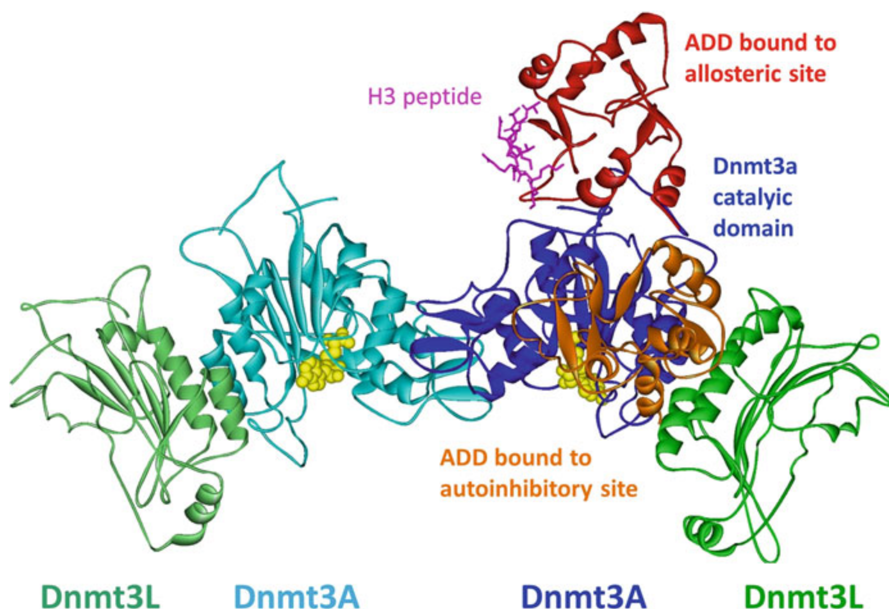


Fig. 4 Structure and allosteric regulation of DNMT3A. The picture shows the structure of the DNMT3A/DNMT3L heterotetramer (Jia et al. 2007). The ADD domain of the *dark blue* DNMT3A subunit is shown in the autoinhibitory conformation (*orange*) and in the active conformation (*red*) (Guo et al. 2015); the ADD domain of the cyan DNMT3A subunit has been omitted for clarity. Binding of the H3 peptide (*purple*) to the ADD domain occurs by interaction with the residues, which are involved in the autoinhibitory-binding interface. Therefore, peptide binding is only possible in the active conformation, and this conformation is consequently stabilized in the presence of the H3 peptide (Guo et al. 2015; Li et al. 2011). (With permission from Jeltsch and Jurkowska 2016, Oxford University Press)

central DNMT3A/DNMT3A interface in the tetramer creates the DNA-binding site, while both interfaces are essential for AdoMet binding and catalytic activity (Jurkowska et al. 2008). The dimerization of DNMT3A via the RD interface increases the size of the DNA interface, and this may compensate for the small TRD of DNMT3A. It is interesting to note that this arrangement is different in prokaryotic DNA MTases, some of which also dimerize (including M.RsrI (Scavetta et al. 2000) and M.MboII (Osipiuk et al. 2003), which have been structurally characterized). Different from DNMT3A, in these cases dimers containing two symmetrically related separate DNA-binding sites are formed.

Recently, new structures of a longer DNMT3A C-terminal fragment including the ADD domain in complex with DNMT3L have been solved and made seminal contributions to our understanding of the mechanism of this enzyme. They showed that the ADD domain can bind to the catalytic domain at two distinct sites, creating two alternative conformations. ADD binding activates the enzyme in one conformation (allosteric binding), while it blocks access to the active center and inhibits catalysis in the other (autoinhibitory binding) (Guo et al. 2015) (Fig. 4). These data indicate

that DNMT3A, like DNMT1, is under tight allosteric control by domain rearrangements, illustrating a fascinating convergence of the regulatory mechanisms of these two enzymes. Similarly as in DNMT1, protein partners can influence the equilibrium of the active and inactive conformations, as it was shown that the stimulatory effect of H3 on DNMT3A depends on its binding to the ADD domain, leading to the stabilization of the ADD at the allosteric binding site (Li et al. 2011; Guo et al. 2015).

4.3 Specificity of DNMT3 Enzymes

Consistent with their designation as de novo MTases, DNMT3A and DNMT3B do not display any significant preference between hemimethylated and unmethylated DNA. Although both enzymes methylate cytosine residues predominantly in the context of CpG dinucleotides, they can also introduce methylation in a non-CpG context (CA >>CT>CC) (Gowher and Jeltsch 2001; Aoki et al. 2001; Ramsahoye et al. 2000). Consistently, methylated non-CpG sites are widespread in embryonic stem (ES) cells and brain, where DNMT3A and DNMT3B enzymes are highly expressed, but not in cells where DNMT3 enzymes are downregulated (Lister et al. 2009; Varley et al. 2013; Lister et al. 2013; Guo et al. 2014). Recently, a survey of the human body epigenome revealed an unexpected presence of non-CpG methylation at lower levels in almost all human tissues (Schultz et al. 2015). Studies with different DNMT KO cell lines confirmed that DNMT3 enzymes introduce the non-CpG methylation in cells (Ziller et al. 2011; Arand et al. 2012). Methylation outside of the CpG context cannot be maintained by DNMT1, which is very specific for CpG sites. Recently, evidence was provided that non-CpG methylation can recruit MeCP2 (Guo et al. 2014; Gabel et al. 2015; Chen et al. 2015), disruption of which is implicated in the Rett syndrome, leading to the repression of long genes in the brain (Gabel et al. 2015).

In addition to their preference for the methylation of CpG sites, both DNMT3A and DNMT3B are very sensitive to the sequences flanking their target sites. This is illustrated by the finding that CpG sites in certain flanking sequences cannot be methylated by DNMT3A at all (Jurkowska et al. 2011c). It was shown that purine bases are preferred at the 5′ end of the CpG sites, whereas pyrimidines are favored at their 3′ end (Lin et al. 2002; Handa and Jeltsch 2005; Jurkowska et al. 2011c). Interestingly, experimental flanking preferences of DNMT3A and DNMT3B correlate with the statistical data on the methylation levels of CpG sites found in the human genome (Handa and Jeltsch 2005; Zhang et al. 2009), suggesting that the inherent sequence preferences of de novo enzymes might contribute to the selection of specific genomic regions that undergo methylation. One further consequence of the strong flanking sequence preferences of DNMT3A and DNMT3B is that both DNA strands of a CpG site, which are embedded in an asymmetric flanking sequence context, usually differ strongly in their preference for DNMT3 methylation. This automatically leads to the preferential methylation of one strand, meaning that DNMT3 enzymes tend to generate hemimethylated products. In vitro experiments have shown that the products of DNMT3A methylation can be readily methylated

by DNMT1 and that both enzymes can act synergistically in the efficient de novo methylation of unmethylated DNA (Fatemi et al. 2002). Mutational analysis of residues in the DNA-binding site of DNMT3A showed that exchanges of critical residues causes massive changes of flanking sequence preferences (Gowher et al. 2006). Interestingly, this includes the exchange at R881, a residue frequently mutated in AML cancer (Hamidi et al. 2015). However, mechanistic understanding of the non-CpG recognition and flanking sequence preferences of DNMT3 enzymes awaits the availability of structures of DNMT3 enzymes with bound substrate DNA.

4.4 Processivity of DNMT3 Enzymes

Initial studies with the C-terminal domains of DNMT3A and DNMT3B showed an interesting difference in the catalytic mechanism of both enzymes. Whereas DNMT3B was shown to methylate multiple CpG sites in a processive manner, DNMT3A was distributive (Gowher and Jeltsch 2002). Later, Reich and colleagues reported that DNMT3A methylates DNA in a processive manner (Holz-Schietinger and Reich 2010). However, at the same time, DNMT3A was shown to bind cooperatively to DNA forming large multimeric protein/DNA fibers (Jia et al. 2007; Jurkowska et al. 2008; Rajavelu et al. 2012) (Fig. 5). These properties appear mutually exclusive, because the concept of a processive turnover is based on isolated enzyme complexes moving along a DNA substrate, which is not compatible with protein complexes multimerizing on DNA. Recent biochemical studies did not detect processive DNA methylation by DNMT3A (Emperle et al. 2014).

4.5 Oligomerization of DNMT3 Enzymes

DNMT3 enzymes have been shown to exhibit a complex oligomerization and multimerization potential including two independent orthogonal multimerization reactions [for a review cf. Jeltsch and Jurkowska 2013]. First, DNMT3A multimerizes on DNA and binds to DNA in a cooperative manner, and second, it can form protein oligomers able to bind to more than one DNA molecule. These two processes will be further described in the next subchapters (Fig. 5).

4.5.1 Multimerization of DNMT3A and DNMT3A/DNMT3L on DNA

As described above, DNMT3A forms a linear heterotetrameric complex with DNMT3L, in which two central DNMT3A subunits interacting via the RD interface generate the DNA-binding pocket (Fig. 5a) (Jia et al. 2007; Jurkowska et al. 2008). DNA binding by DNMT3A is nonspecific (Rajavelu et al. 2012), and DNMT3A (and DNMT3A/DNMT3L) complexes polymerize on DNA by binding next to each other and forming DNMT3A-DNA filaments (Jurkowska et al. 2008; Rajavelu et al. 2012) (Fig. 5). Interestingly, two adjacent DNMT3 complexes in such filament contact one CpG site in both DNA strands providing an option for the enzyme to directly methylate both DNA strands (Jurkowska et al. 2008).

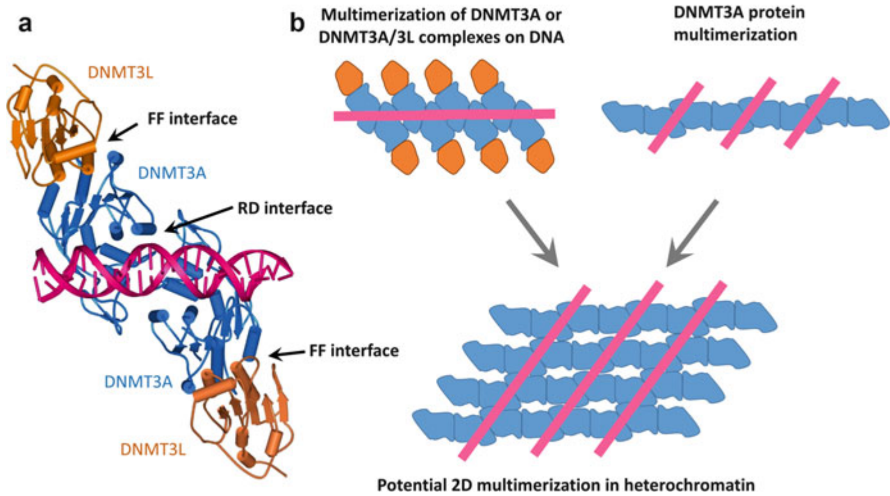


Fig. 5 Multimerization of DNMT3A and DNMT3A/DNMT3L complexes. (**a**) Structure of the DNMT3A/DNMT3L complex with bound DNA is shown. (**b**) Schematic models of DNMT3A multimerization on DNA, protein multimerization and binding to several DNA molecules, and the combination of both processes

However, a productive interaction with neighboring CpG sites is only possible if they are present in a distance of approximately 10 bps, due to the spacing of the two active centers at the RD interface of the individual DNMT3 complex subunits. Indeed, in vitro methylation experiments demonstrated that there is a correlation of methylation between sites localized ~10 bps apart (Jia et al. 2007; Jurkowska et al. 2008). Interestingly, an enrichment of CpG sites in such distance is observed in the differentially methylated regions (DMRs) of 12 maternally imprinted mouse genes, which are biological substrates of the DNMT3A/DNMT3L complex, suggesting that the favorable CpG spacing could make these sequences good substrates for the MTase complex (Jia et al. 2007). Of note, the 10 bps periodicity of the methylation patterns has been observed in genome-wide methylation studies (Lister et al. 2009; Smallwood et al. 2011).

Multimerization of DNMT3A or DNMT3A/DNMT3L tetramers on DNA leads to a cooperative DNA binding, as confirmed by different methods, including cooperative binding detected in gel retardation assays, sigmoidal binding curves of DNA substrates observed in solution DNA-binding experiments, and direct imaging of DNMT3-DNA filaments by atomic force microscopy (Jia et al. 2007; Jurkowska et al. 2008; Rajavelu et al. 2012; Emperle et al. 2014). The interface of adjacent DNMT3A complexes bound to DNA has been mapped to a loop within the putative TRD of DNMT3A, and mutation of residues within this region disrupted multimerization (Rajavelu et al. 2012). Interestingly, it also led to the loss of heterochromatic enrichment of DNMT3A, suggesting that cooperative DNA binding and multimerization of DNMT3A complexes on DNA contribute to the heterochromatic localization of the enzyme in cells. Recent studies have further shown that the cooperative

binding of DNMT3A to long DNA substrates increases the rate of DNA methylation (Emperle et al. 2014), indicating that it is important for DNA methylation by DNMT3A. However, the exact role of cooperative DNA binging of DNMT3A in cells needs further investigation. In addition, the sizes of DNMT3A-DNA filaments in living cells are currently unknown; one may speculate that binding of up to five complexes would be possible in the linker DNA regions. This is in agreement with biochemical data showing preferential methylation of linker DNA by DNMT3 enzymes in vitro (Gowher et al. 2005b; Takeshima et al. 2008; Felle et al. 2011). In vivo studies also confirmed this observation, showing that DNMT3B expressed in yeast preferentially methylates linker DNA (Morselli et al. 2015). A similar pattern was also observed after the reintroduction of DNMT3 enzymes into corresponding KO ES cell lines (Baubec et al. 2015). Longer filaments may form if DNMT3 binding is coupled with nucleosome remodeling. Consistently, DNMT3 enzymes were shown to form complexes with various chromatin remodelers, including SMARCA4 (Datta et al. 2005), CHD4 (Cai et al. 2014), hSNF2 (Geiman et al. 2004), and HELLS (Zhu et al. 2006; Myant and Stancheva 2008), and the interaction with HELLS is essential for DNA methylation (Muegge 2005). In line with this model, it has been documented that remodeling activity promotes the methylation of nucleosomal DNA (Felle et al. 2011). One important functional aspect of the cooperative DNA binding of DNMT3A may be that it increases the DNA-binding affinity and reduces the rate of dissociation, which may help to anchor the MTase on DNA, in agreement with its strong binding to methylated chromatin (Jeong et al. 2009; Sharma et al. 2011).

4.5.2 Protein Multimerization of DNMT3 Enzymes

After the discovery of the heterotetrameric structure of DNMT3A/DNMT3L, it was also shown that DNMT3A alone forms protein filaments (Fig. 5b), which can lead to its reversible aggregation as observed in different studies (Jurkowska et al. 2011b; Kareta et al. 2006). The reason for this is that the FF interface of the DNMT3A/DNMT3L tetramer is symmetric, so that it also supports the symmetric interaction of two DNMT3A molecules in addition to the mixed interaction of DNMT3A with DNMT3L. Hence, each DNMT3A subunit contains two interfaces for trophic interaction with another DNMT3A subunit, the RD interface and the FF interface, which explains why it can form protein fibers. Notably, the addition of DNMT3L directs the preferential formation of defined DNMT3A/DNMT3L heterotetramers that cannot oligomerize further, because DNMT3L does not contain an RD interface and therefore functions as a cap in protein multimerization. As described below, this process has been implicated in the release of DNMT3A from heterochromatic sites by the addition of DNMT3L (Jurkowska et al. 2011b).

Since each RD interface of such multimeric DNMT3A oligomer constitutes a potential DNA-binding site, the protein oligomers can bind to more than one DNA molecule, provided that they are oriented roughly in parallel, as shown by biophysical experiments (Jurkowska et al. 2011b). The ability to form protein oligomers plays a central role in the heterochromatin localization of DNMT3A, as non-oligomerizing DNMT3A mutants affected at the interfaces lost the ability to bind to

heterochromatin, despite the presence of intact PWWP and ADD domains. Since heterochromatic DNA is densely packed, it can provide several DNA strands for DNMT3A interaction in matching geometry, and this might contribute to guiding DNMT3A to pericentromeric chromatin.

Despite significant progress in dissecting protein multimerization of the DNMT3 enzymes, many questions are still not resolved. For example, DNMT3A has been shown to form catalytically active heterodimers with DNMT3B that use the same interfaces as described above for DNMT3A (Li et al. 2007). However, it is currently unknown what are the relative affinities for the symmetric DNMT3A and DNMT3B interactions, as compared to the mixed interaction of DNMT3A and DNMT3B at the two interfaces. Moreover, the relative preferences for binding DNMT3L at the FF interface are also unknown, although the formation of defined heterotetramers of DNMT3A and DNMT3L suggests that the DNMT3A/DNMT3L interaction is preferred over the DNMT3A/DNMT3A interaction. Finally, the direct proof for the existence of DNMT3 protein multimers in cells that are larger than the tetrameric structure observed in the DNMT3A/DNMT3L complex still needs to be provided.

4.6 Direct Chromatin Interaction of DNMT3 Enzymes

4.6.1 Binding of the DNMT3 ADD Domain to H3 Tails

The ADD domains of all three DNMT3 proteins were shown to interact specifically with histone H3 tails unmethylated at lysine 4 (Fig. 4), and the binding was disrupted by either the di- and trimethylation or acetylation of K4 or the acetylation of the N-terminus of H3 (Ooi et al. 2007; Otani et al. 2009; Zhang et al. 2010; Noh et al. 2015). Interestingly, H3K4me1, which is observed at enhancers, does not hinder the binding of ADD much, but phosphorylation of T6 does (Zhang et al. 2010; Noh et al. 2015). The structures of the ADD domains from DNMT3A and DNMT3L in complex with histone H3 tail peptides have been solved (Ooi et al. 2007; Otani et al. 2009). Interestingly, binding to H3 tails stimulates the methylation of chromatin-bound DNA by DNMT3A in vitro (Zhang et al. 2010) and it directly activates DNMT3A by an allosteric mechanism (Li et al. 2011). This regulatory mechanism has recently been confirmed in a structural analysis by Xu and colleagues, which showed that the ADD domain could bind to the catalytic domain of DNMT3A at two sites, an allosteric site and an autoinhibitory site. H3 peptide binding stabilizes the active conformation, leading to an allosteric activation of DNMT3A (Guo et al. 2015). These results indicate that the ADD domain of DNMT3A can guide DNA methylation in response to specific histone modifications and provide the first evidence that DNA methyltransferases could be targeted to chromatin carrying specific marks. Indeed, a strong correlation of DNA methylation with the absence of H3K4me3 was observed in several genome-wide studies (Hodges et al. 2009; Meissner et al. 2008; Weber et al. 2007; Zhang et al. 2009) suggesting that this mechanism plays an important role in the generation of the genomic DNA methylation pattern. This hypothesis was recently experimentally verified, when it was shown that a DNMT3A enzyme with an engineered ADD domain able to tolerate

K4 methylation or T6 phosphorylation generates abnormal DNA methylation patterns in cells (Noh et al. 2015) and DNMT3B artificially introduced in yeast does not methylate genomic regions with high H3K4me3 content (Morselli et al. 2015).

4.6.2 Binding of DNMT3 PWWP Domain to H3 Methylated at K36

The PWWP domain is essential for the targeting of DNMT3A and DNMT3B to pericentromeric chromatin (Chen et al. 2004; Ge et al. 2004) and to gene bodies, via specific recognition of histone H3 tails trimethylated at lysine 36 (Dhayalan et al. 2010). In addition, the interaction of DNMT3A with H3K36me3 increases the activity of DNMT3A on chromatin, which carries this mark (Dhayalan et al. 2010). However, the molecular mechanism of this regulation remains unknown. These findings can explain the genome-wide correlation of DNA methylation and H3K36me3 methylation both in the gene bodies and in heterochromatin. H3K36me3 accumulates in euchromatin in the body of active genes, and its distribution is anticorrelated with H3K4me3 (Barski et al. 2007; Edmunds et al. 2008; Guenther et al. 2007; Larschan et al. 2007; Vakoc et al. 2006). DNA methylation of gene bodies mirrors that pattern, with gene bodies of active genes showing high and those of inactive genes low methylation (Ball et al. 2009; Hellman and Chess 2007). Additionally, a correlation between H3K36me3 and DNA methylation was observed at exon-intron boundaries, with exons showing increased levels of both H3K36me3 (Kolasinska-Zwierz et al. 2009) and DNA methylation (Hodges et al. 2009). A subset of heterochromatin containing repetitive sequences with copy number variations is strongly enriched in H3K36me3 (Ernst et al. 2011), which can explain the role of the DNMT3A PWWP domain in the heterochromatic localization of the enzyme and the strong correlation of DNA methylation and H3K36me3 observed in genome-wide DNA methylation studies (Meissner et al. 2008; Hodges et al. 2009). The central role of K36 methylation in targeting of DNA methylation has been experimentally confirmed in yeast (Morselli et al. 2015) and in a study showing that the methylation of gene bodies by DNMT3B directly depends on H3K36 methylation and an intact DNMT3B PWWP domain (Baubec et al. 2015).

In addition to H3 binding, the PWWP domains of DNMT3A and DNMT3B were shown to interact with DNA to a variable degree, with DNMT3B PWWP binding DNA more strongly (Qiu et al. 2002; Purdy et al. 2010). Similarly, a combined interaction with methylated H3K36 and DNA has been observed for other PWWP domains as well (Dhayalan et al. 2010; van Nuland et al. 2013). This finding is not unexpected, as the K36 side chain emerges from the nucleosome body next to one turn of the bound DNA. H3K36me3 and DNA binding by PWWP domains are mediated by two adjacent interfaces, one featuring an aromatic cage for peptide binding and the other one displaying a basic region for DNA interaction. Recently, a mechanism for the recognition of nucleosomes and DNA methylation by DNMT3A has been proposed, based on a structural model of the full-length DNMT3A/DNMT3L heterotetramer in complex with an H3K36me3-modified dinucleosome (Rondelet et al. 2016). It suggested that the targeting of DNMT3A occurs through a specific recognition and binding of H3K36me3 by the PWWP domain, which is followed by an activation of the catalytic domain through the

binding of H3 tails unmodified at K4 to the ADD domain, resulting in the methylation of nearby cytosines. This model predicts that DNA methylation by the heterotetramer would occur on the linker DNA between the nucleosomes, which is in agreement with results of in vitro and in vivo studies as described above (Gowher et al. 2005b; Takeshima et al. 2008; Felle et al. 2011; Baubec et al. 2015; Morselli et al. 2015). Overall, the multivalent interaction of the DNMT3 enzymes with chromatin through multimerization and the ADD and PWWP domains may explain the extraordinarily strong binding of these enzymes to nucleosomal heterochromatic DNA (Jeong et al. 2009; Sharma et al. 2011), which consequently is methylated in the cell.

4.7 DNMT3-Interacting Proteins

Up to date, the interaction of DNMT3 enzymes with DNMT3L and MeCP2 has been studied in detail, revealing important roles in targeting, allosteric regulation, and control of multimerization. Unfortunately, for most other DNMT3-interacting proteins, detailed information about their function is not yet available.

4.7.1 DNMT3A/DNMT3L Interaction

DNMT3L co-localizes with both DNMT3A and DNMT3B in mammalian cells (Hata et al. 2002). It directly interacts with its C-terminal domain with the catalytic domains of DNMT3A and DNMT3B and stimulates the activity of both enzymes in vivo (Chedin et al. 2002; Chen et al. 2005) and in vitro (Suetake et al. 2004; Gowher et al. 2005a; Kareta et al. 2006). DNMT3L is expressed during gametogenesis and embryonic stages of development (Bourc'his et al. 2001; Hata et al. 2002; Bourc'his and Bestor 2004), where it functions as a stimulatory factor of DNMT3A and is needed to establish DNA methylation patterns in the developing germ line cells. The structure of the complex of the C-terminal domains of DNMT3A and DNMT3L provided the structural basis for the DNMT3A/DNMT3L interaction and offered mechanistic explanation for the observed stimulatory effect of DNMT3L. It revealed that the interaction of DNMT3A with DNMT3L through the FF interface influences the structure of DNMT3A via the α-helices C, D, and E. Residues from these helices directly interact with the key catalytic and AdoMet-binding residues, which may explain the stimulatory effect DNMT3L exerts on DNMT3A AdoMet binding and catalysis (Jia et al. 2007).

As described above, binding of DNMT3L to DNMT3A leads to the disruption of DNMT3A protein oligomers, and this changes the subnuclear localization of DNMT3A in cells (Fig. 6). In vivo, DNMT3L was shown to release DNMT3A from heterochromatin, by disrupting large DNMT3A oligomers and converting them into defined tetramers, which are homogenously distributed in the cell nucleus (Jurkowska et al. 2011b). The redistribution of DNMT3A may be important for the methylation of imprinted differentially methylated regions (DMRs) and other targets in gene promoters, which generally are euchromatic. This finding goes in line with the discovery that DNMT3L favors DNA methylation in gene bodies (Neri

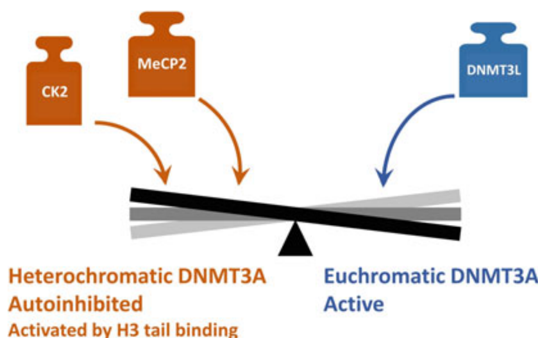


Fig. 6 Mechanisms regulating the activity and localization of DNMT3A. Different interactors and PTMs regulate the activity and localization in a concerted fashion. DNMT3L stimulates DNMT3A and promotes its euchromatic localization. Contrarily, MeCP2 interaction and CK2-mediated phosphorylation downregulate the activity of DNMT3A and promote its heterochromatic localization, where the interaction with modified H3 tails could allosterically stimulate the enzyme

et al. 2013). Hence DNMT3L, which was originally discovered as a stimulator of DNMT3A (Gowher et al. 2005a), also changes the subnuclear localization of this enzyme (Jurkowska et al. 2011b). Recent data indicate that the combined regulation of activity and localization of DNMT3A also applies to other regulatory cues (see below for MeCP2 interaction and CK2-mediated phosphorylation of DNMT3A) and might be a general mechanism of regulation for this family of enzymes (Figs. 3 and 4).

4.7.2 Interaction of DNMT3A with MeCP2

Recently, we identified the chromosomal protein MeCP2, which binds methylated DNA with its methyl-binding domain (MBD), as a direct and strong interactor of DNMT3A and DNMT3B, and mapped this interaction to the TRD of MeCP2 and the ADD domain of the DNMT3 enzymes (Rajavelu et al., manuscript in preparation). Binding of MeCP2 resulted in a strong reduction of the DNA methylation activity of DNMT3A in vitro, and overexpression of MeCP2 in human cells led to a global reduction of DNA methylation. We could show that binding of MeCP2 allosterically stabilizes the autoinhibitory conformation of DNMT3A. Interestingly, this interaction and its resulting inhibition were relieved by histone H3 binding to DNMT3A. In addition, we also observed that MeCP2 contributes to the heterochromatic targeting of DNMT3A. These findings led to a model of an allosteric control of the target site specificity of DNMT3A by the combined effects of interacting partners, like MeCP2 and histone H3 tails. In this model, MeCP2 binding inactivates DNMT3A, thereby preventing aberrant methylation of bulk DNA. At the same time, it helps to deliver DNMT3A to heterochromatin. After binding to chromatin, which presents H3 tails modified in a PTM pattern matching the specificity of DNMT3A, the allosteric inhibition is relieved, and the activated enzyme can methylate its target sites.

Interestingly, MeCP2 functions as a perfect antagonist of DNMT3L, which increases the activity of DNMT3A and leads to its release from heterochromatin as described above. Moreover, regulation of DNMT3A by CK2 (which is described below) resembles the MeCP2 effect, since it reduces the activity of DNMT3A and contributes to the heterochromatic sequestering of the methyltransferase. This illustrates an unexpected mechanistic convergence in the regulation and targeting of DNMT3A by interactors and posttranslational modifications (Fig. 6).

4.8 Phosphorylation of DNMT3A

The regulation of the DNMT3 enzymes by phosphorylation has not been studied almost at all, despite the fact that >70 phosphorylation sites have been identified in DNMT3A and DNMT3B in global proteomic studies (http://www.phosphosite.org). The unique example has been recently provided investigates the phosphorylation of DNMT3A by Casein Kinase 2 (CK2) (Deplus et al. 2014). CK2 is a so-called survival protein kinase, which suppresses cancer cell death and is often upregulated in cancers. It was shown that CK2 phosphorylates DNMT3A at two sites, S386 and S389, located next to the PWWP domain, and that CK2-mediated phosphorylation increases the heterochromatic targeting of DNMT3A and reduces its DNA methylation activity (Deplus et al. 2014). This effect was reflected by changes in the cellular DNA methylation after CK2 knockout, which may explain global hypomethylation in cancer cells overexpressing CK2. These data further support the view that the combined regulation of enzymatic activity and localization is a general principle in the regulation of DNMT3A (as already observed for the DNMT3L and MeCP2 interaction) (Fig. 6).

4.9 Binding of Regulatory DNA and RNA to DNMT3 Enzymes

Similar as in DNMT1, additional DNA-binding sites have been identified in the N-terminal part of DNMT3 enzymes. As described above, the isolated PWWP domain of DNMT3B has a DNA-binding activity (Qiu et al. 2002). In DNMT3A, an additional DNA-binding site was detected and connected to the PWWP domain as well (Purdy et al. 2010). Moreover, the very N-terminal part of DNMT3A was shown to bind DNA (Suetake et al. 2011). Later, it was shown that long noncoding RNAs bind strongly to the catalytic domain of DNMT3A, causing inhibition of the enzyme (Holz-Schietinger and Reich 2012). In addition, the authors detected binding of RNA to allosteric sites, which did not change the catalytic activity. Besides, it was shown that a non-coding RNA derived from the rDNA promoter binds to the promoter forming RNA/DNA triplex structures that are specifically recognized by DNMT3B, establishing a novel pathway of RNA-directed DNA methylation (Schmitz et al. 2010; Bierhoff et al. 2010). Future work will show if RNA triplex-based recruitment emerges as new and general principle of the RNA-dependent recruitment of DNMTs and other chromatin-interacting enzymes.

5 Outlook

After almost 40 years of intensive research in the DNA methylation field, we have learned a great deal about the biochemical, structural, and enzymatic properties of the mammalian DNA methyltransferases. However, their regulation in cells has only begun to be uncovered. Importantly, it has been lately realized that the precise control of DNMT activity is critically involved in the generation and maintenance of the dynamic DNA methylation patterns in living cells. Recent crystallographic studies with DNMT1 and DNMT3A revealed that both enzymes unexpectedly undergo large domain rearrangements, which allosterically regulate their catalytic activity. This unforeseen discovery leads to the important conclusion that by influencing domain rearrangements, any posttranslational modifications and interaction partners, (like proteins, allosteric DNA, or noncoding RNAs) at various parts of the methyltransferases could directly regulate the enzymatic activity and specificity of the DNMTs via allosteric effects, providing new and fascinating perspectives on the investigation of the effects of interactors and PTMs on these enzymes. Finally, the connection of DNMTs and non-coding RNA adds a novel and yet underexplored link between the sequence of the genome and its DNA methylation pattern.

References

Allis CD, Caparos M-L, Jenuwein T, Reinberg D. Epigenetics. 2nd ed. Woodbury: CSH Laboratory Press; 2015.

Aoki A, Suetake I, Miyagawa J, Fujio T, Chijiwa T, Sasaki H, et al. Enzymatic properties of de novo-type mouse DNA (cytosine-5) methyltransferases. Nucleic Acids Res. 2001;29(17):3506–12.

Arand J, Spieler D, Karius T, Branco MR, Meilinger D, Meissner A, et al. In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. PLoS Genet. 2012;8(6):e1002750.

Arita K, Ariyoshi M, Tochio H, Nakamura Y, Shirakawa M. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature. 2008;455(7214):818–21.

Armstrong L. Epigenetics. New York: Garland Science; 2013.

Avvakumov GV, Walker JR, Xue S, Li Y, Duan S, Bronner C, et al. Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. Nature. 2008;455(7214):822–5.

Bacolla A, Pradhan S, Roberts RJ, Wells RD. Recombinant human DNA (cytosine-5) methyltransferase. II. Steady-state kinetics reveal allosteric activation by methylated dna. J Biol Chem. 1999;274(46):33011–9.

Bacolla A, Pradhan S, Larson JE, Roberts RJ, Wells RD. Recombinant human DNA (cytosine-5) methyltransferase. III. Allosteric control, reaction order, and influence of plasmid topology and triplet repeat length on methylation of the fragile X CGG.CCG sequence. J Biol Chem. 2001;276(21):18605–13.

Ball MP, Li JB, Gao Y, Lee JH, LeProust EM, Park IH, et al. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nat Biotechnol. 2009;27(4):361–8.

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. Cell. 2007;129(4):823–37.

Bashtrykov P, Jankevicius G, Smarandache A, Jurkowska RZ, Ragozin S, Jeltsch A. Specificity of Dnmt1 for methylation of hemimethylated CpG sites resides in its catalytic domain. Chem Biol. 2012a;19(5):572–8.

Bashtrykov P, Ragozin S, Jeltsch A. Mechanistic details of the DNA recognition by the Dnmt1 DNA methyltransferase. FEBS Lett. 2012b;586(13):1821–3.
Bashtrykov P, Jankevicius G, Jurkowska RZ, Ragozin S, Jeltsch A. The UHRF1 protein stimulates the activity and specificity of the maintenance DNA methyltransferase DNMT1 by an allosteric mechanism. J Biol Chem. 2014a;289(7):4106–15.
Bashtrykov P, Rajavelu A, Hackner B, Ragozin S, Carell T, Jeltsch A. Targeted mutagenesis results in an activation of DNA methyltransferase 1 and confirms an autoinhibitory role of its RFTS domain. Chembiochem Eur J Chem Biol. 2014b;15(5):743–8.
Baubec T, Colombo DF, Wirbelauer C, Schmidt J, Burger L, Krebs AR, et al. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature. 2015;520(7546):243–7.
Baylin SB, Jones PA. A decade of exploring the cancer epigenome – biological and translational implications. Nat Rev Cancer. 2011;11(10):726–34.
Bergman Y, Cedar H. DNA methylation dynamics in health and disease. Nat Struct Mol Biol. 2013;20(3):274–81.
Berkyurek AC, Suetake I, Arita K, Takeshita K, Nakagawa A, Shirakawa M, et al. The DNA methyltransferase Dnmt1 directly interacts with the SET and RING finger-associated (SRA) domain of the multifunctional protein Uhrf1 to facilitate accession of the catalytic center to hemimethylated DNA. J Biol Chem. 2014;289(1):379–86.
Bierhoff H, Schmitz K, Maass F, Ye J, Grummt I. Noncoding transcripts in sense and antisense orientation regulate the epigenetic state of ribosomal RNA genes. Cold Spring Harb Symp Quant Biol. 2010;75:357–64.
Bonasio R, Tu S, Reinberg D. Molecular signals of epigenetic states. Science. 2010;330(6004):612–6.
Bostick M, Kim JK, Esteve PO, Clark A, Pradhan S, Jacobsen SE. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science. 2007;317(5845):1760–4.
Bourc'his D, Bestor TH. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature. 2004;431(7004):96–9.
Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH. Dnmt3L and the establishment of maternal genomic imprints. Science. 2001;294(5551):2536–9.
Brenner C, Deplus R, Didelot C, Loriot A, Vire E, De Smet C, et al. Myc represses transcription through recruitment of DNA methyltransferase corepressor. EMBO J. 2005;24(2):336–46.
Cai Y, Geutjes EJ, de Lint K, Roepman P, Bruurs L, Yu LR, et al. The NuRD complex cooperates with DNMTs to maintain silencing of key colorectal tumor suppressor genes. Oncogene. 2014;33(17):2157–68.
Chedin F, Lieber MR, Hsieh CL. The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. Proc Natl Acad Sci U S A. 2002;99(26):16916–21.
Chen TP, Tsujimoto N, Li E. The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. Mol Cell Biol. 2004;24(20):9048–58.
Chen ZX, Mann JR, Hsieh CL, Riggs AD, Chedin F. Physical and functional interactions between the human DNMT3L protein and members of the de novo methyltransferase family. J Cell Biochem. 2005;95(5):902–17.
Chen L, Chen K, Lavery LA, Baker SA, Shaw CA, Li W, et al. MeCP2 binds to non-CG methylated DNA as neurons mature, influencing transcription and the timing of onset for Rett syndrome. Proc Natl Acad Sci U S A. 2015;112(17):5509–14.
Cheng X, Blumenthal RM. Mammalian DNA methyltransferases: a structural perspective. Structure. 2008;16(3):341–50.
Cheng X, Roberts RJ. AdoMet-dependent methylation, DNA methyltransferases and base flipping. Nucleic Acids Res. 2001;29(18):3784–95.
Cheng J, Yang H, Fang J, Ma L, Gong R, Wang P, et al. Molecular mechanism for USP7-mediated DNMT1 stabilization by acetylation. Nat Commun. 2015;6:7023.
Choudhary C, Kumar C, Gnad F, Nielsen ML, Rehman M, Walther TC, et al. Lysine acetylation targets protein complexes and co-regulates major cellular functions. Science. 2009;325(5942):834–40.

Christman JK, Sheikhnejad G, Marasco CJ, Sufrin JR. 5-Methyl-2′-deoxycytidine in single-stranded DNA can act in cis to signal de novo DNA methylation. Proc Natl Acad Sci U S A. 1995;92(16):7347–51.

Chuang LS, Ian HI, Koh TW, Ng HH, Xu G, Li BF. Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. Science. 1997;277(5334):1996–2000.

Datta J, Majumder S, Bai S, Ghoshal K, Kutay H, Smith DS, et al. Physical and functional interaction of DNA methyltransferase 3A with Mbd3 and Brg1 in mouse lymphosarcoma cells. Cancer Res. 2005;65(23):10891–900.

Deplus R, Blanchon L, Rajavelu A, Boukaba A, Defrance M, Luciani J, et al. Regulation of DNA methylation patterns by CK2-mediated phosphorylation of Dnmt3a. Cell Rep. 2014;8(3):743–53.

Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S, et al. The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. J Biol Chem. 2010;285(34):26114–20.

Di Ruscio A, Ebralidze AK, Benoukraf T, Amabile G, Goff LA, Terragni J, et al. DNMT1-interacting RNAs block gene-specific DNA methylation. Nature. 2013;503(7476):371–6.

Du Z, Song J, Wang Y, Zhao Y, Guda K, Yang S, et al. DNMT1 stability is regulated by proteins coordinating deubiquitination and acetylation-driven ubiquitination. Sci Signal. 2010;3(146):ra80.

Du Q, Wang Z, Schramm VL. Human DNMT1 transition state structure. Proc Natl Acad Sci U S A. 2016;113(11):2916–21.

Easwaran HP, Schermelleh L, Leonhardt H, Cardoso MC. Replication-independent chromatin loading of Dnmt1 during G2 and M phases. EMBO Rep. 2004;5(12):1181–6.

Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, et al. DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet. 2006;38(12):1378–85.

Edmunds JW, Mahadevan LC, Clayton AL. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. EMBO J. 2008;27(2):406–20.

Egger G, Jeong S, Escobar SG, Cortez CC, Li TW, Saito Y, et al. Identification of DNMT1 (DNA methyltransferase 1) hypomorphs in somatic knockouts suggests an essential role for DNMT1 in cell survival. Proc Natl Acad Sci U S A. 2006;103(38):14080–5.

Emperle M, Rajavelu A, Reinhardt R, Jurkowska RZ, Jeltsch A. Cooperative DNA binding and protein/DNA fiber formation increases the activity of the Dnmt3a DNA methyltransferase. J Biol Chem. 2014;289(43):29602–13.

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011;473(7345):43–9.

Esteve PO, Chin HG, Benner J, Feehery GR, Samaranayake M, Horwitz GA, et al. Regulation of DNMT1 stability through SET7-mediated lysine methylation in mammalian cells. Proc Natl Acad Sci U S A. 2009;106(13):5076–81.

Esteve PO, Chang Y, Samaranayake M, Upadhyay AK, Horton JR, Feehery GR, et al. A methylation and phosphorylation switch between an adjacent lysine and serine determines human DNMT1 stability. Nat Struct Mol Biol. 2011;18(1):42–8.

Esteve PO, Zhang G, Ponnaluri VK, Deepti K, Chin HG, Dai N, et al. Binding of 14-3-3 reader proteins to phosphorylated DNMT1 facilitates aberrant DNA methylation and gene expression. Nucleic Acids Res. 2016;44(4):1642–56.

Fatemi M, Hermann A, Pradhan S, Jeltsch A. The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. J Mol Biol. 2001;309(5):1189–99.

Fatemi M, Hermann A, Gowher H, Jeltsch A. Dnmt3a and Dnmt1 functionally cooperate during de novo methylation of DNA. Eur J Biochemistry FEBS. 2002;269(20):4981–4.

Felle M, Hoffmeister H, Rothammer J, Fuchs A, Exler JH, Langst G. Nucleosomes protect DNA from DNA methylation in vivo and in vitro. Nucleic Acids Res. 2011;39(16):6956–69.

Flynn J, Fang JY, Mikovits JA, Reich NO. A potent cell-active allosteric inhibitor of murine DNA cytosine C5 methyltransferase. J Biol Chem. 2003;278(10):8238–43.

Fuks F, Burgers WA, Godin N, Kasai M, Kouzarides T. Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. EMBO J. 2001;20(10):2536–44.
Fuks F, Hurd PJ, Deplus R, Kouzarides T. The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. Nucleic Acids Res. 2003a;31(9):2305–12.
Fuks F, Hurd PJ, Wolf D, Nan X, Bird AP, Kouzarides T. The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. J Biol Chem. 2003b;278(6):4035–40.
Gabel HW, Kinde B, Stroud H, Gilbert CS, Harmin DA, Kastan NR, et al. Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. Nature. 2015;522(7554):89–93.
Ge YZ, Pu MT, Gowher H, Wu HP, Ding JP, Jeltsch A, et al. Chromatin targeting of de novo DNA methyltransferases by the PWWP domain. J Biol Chem. 2004;279(24):25447–54.
Geiman TM, Sankpal UT, Robertson AK, Zhao Y, Robertson KD. DNMT3B interacts with hSNF2H chromatin remodeling enzyme, HDACs 1 and 2, and components of the histone methylation system. Biochem Biophys Res Commun. 2004;318(2):544–55.
Glickman JF, Flynn J, Reich NO. Purification and characterization of recombinant baculovirus-expressed mouse DNA methyltransferase. Biochem Biophys Res Commun. 1997a;230(2):280–4.
Glickman JF, Pavlovich JG, Reich NO. Peptide mapping of the murine DNA methyltransferase reveals a major phosphorylation site and the start of translation. J Biol Chem. 1997b;272(28):17851–7.
Gowher H, Jeltsch A. Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpA sites. J Mol Biol. 2001;309(5):1201–8.
Gowher H, Jeltsch A. Molecular enzymology of the catalytic domains of the Dnmt3a and Dnmt3b DNA methyltransferases. J Biol Chem. 2002;277(23):20409–14.
Gowher H, Liebert K, Hermann A, Xu G, Jeltsch A. Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. J Biol Chem. 2005a;280(14):13341–8.
Gowher H, Stockdale CJ, Goyal R, Ferreira H, Owen-Hughes T, Jeltsch A. De novo methylation of nucleosomal DNA by the mammalian Dnmt1 and Dnmt3A DNA methyltransferases. Biochemistry. 2005b;44(29):9899–904.
Gowher H, Loutchanwoot P, Vorobjeva O, Handa V, Jurkowska RZ, Jurkowski TP, et al. Mutational analysis of the catalytic domain of the murine Dnmt3a DNA-(cytosine C5)-methyltransferase. J Mol Biol. 2006;357(3):928–41.
Goyal R, Reinhardt R, Jeltsch A. Accuracy of DNA methylation pattern preservation by the Dnmt1 methyltransferase. Nucleic Acids Res. 2006;34(4):1182–8.
Goyal R, Rathert P, Laser H, Gowher H, Jeltsch A. Phosphorylation of serine-515 activates the Mammalian maintenance methyltransferase Dnmt1. Epigenetics. 2007;2(3):155–60.
Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. A chromatin landmark and transcription initiation at most promoters in human cells. Cell. 2007;130(1):77–88.
Guo JU, Su Y, Shin JH, Shin J, Li H, Xie B, et al. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. Nat Neurosci. 2014;17(2):215–22.
Guo X, Wang L, Li J, Ding Z, Xiao J, Yin X, et al. Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature. 2015;517(7536):640–4.
Hamidi T, Singh AK, Chen T. Genetic alterations of DNA methylation machinery in human diseases. Epigenomics. 2015;7(2):247–65.
Handa V, Jeltsch A. Profound flanking sequence preference of Dnmt3a and Dnmt3b mammalian DNA methyltransferases shape the human epigenome. J Mol Biol. 2005;348(5):1103–12.
Hashimoto H, Horton JR, Zhang X, Bostick M, Jacobsen SE, Cheng X. The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. Nature. 2008;455(7214):826–9.
Hata K, Okano M, Lei H, Li E. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Development. 2002;129(8):1983–93.
Hellman A, Chess A. Gene body-specific methylation on the active X chromosome. Science. 2007;315(5815):1141–3.

Hermann A, Gowher H, Jeltsch A. Biochemistry and biology of mammalian DNA methyltransferases. Cell Mole Life Sci CMLS. 2004a;61(19–20):2571–87.

Hermann A, Goyal R, Jeltsch A. The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. J Biol Chem. 2004b;279(46):48350–9.

Hervouet E, Lalier L, Debien E, Cheray M, Geairon A, Rogniaux H, et al. Disruption of Dnmt1/PCNA/UHRF1 interactions promotes tumorigenesis from human and mice glial cells. PLoS One. 2010;5(6):e11333.

Hodges E, Smith AD, Kendall J, Xuan Z, Ravi K, Rooks M, et al. High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. Genome Res. 2009;19(9):1593–605.

Holoch D, Moazed D. RNA-mediated epigenetic regulation of gene expression. Nat Rev Genet. 2015;16(2):71–84.

Holz-Schietinger C, Reich NO. The inherent processivity of the human de novo methyltransferase 3A (DNMT3A) is enhanced by DNMT3L. J Biol Chem. 2010;285(38):29091–100.

Holz-Schietinger C, Reich NO. RNA modulation of the human DNA methyltransferase 3A. Nucleic Acids Res. 2012;40(17):8550–7.

Hu L, Li Z, Wang P, Lin Y, Xu Y. Crystal structure of PHD domain of UHRF1 and insights into recognition of unmodified histone H3 arginine residue 2. Cell Res. 2011;21(9):1374–8.

Iida T, Suetake I, Tajima S, Morioka H, Ohta S, Obuse C, et al. PCNA clamp facilitates action of DNA cytosine methyltransferase 1 on hemimethylated DNA. Genes Cells Devoted Mole Cell Mech. 2002;7(10):997–1007.

Iwasaki YW, Siomi MC, Siomi H. PIWI-interacting RNA: its biogenesis and functions. Annu Rev Biochem. 2015;84:405–33.

Jeltsch A. Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. Chem Eur J Chem Biol. 2002;3(4):275–93.

Jeltsch A. On the enzymatic properties of Dnmt1: specificity, processivity, mechanism of linear diffusion and allosteric regulation of the enzyme. Epigenetics. 2006;1(2):63–6.

Jeltsch A. Reading and writing DNA methylation. Nat Struct Mol Biol. 2008;15(10):1003–4.

Jeltsch A, Jurkowska RZ. Multimerization of the dnmt3a DNA methyltransferase and its functional implications. Prog Mol Biol Transl Sci. 2013;117:445–64.

Jeltsch A, Jurkowska RZ. New concepts in DNA methylation. Trends Biochem Sci. 2014;39(7):310–8.

Jeltsch A, Jurkowska RZ. Allosteric control of mammalian DNA Methyltransferases – a new regulatory paradigm. Nucl. Acids Res. 2016. doi:10.1093/nar/gkw723.

Jeong S, Liang GN, Sharma S, Lin JC, Choi SH, Han H, et al. Selective anchoring of DNA methyltransferases 3A and 3B to nucleosomes containing methylated DNA. Mol Cell Biol. 2009;29(19):5366–76.

Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature. 2007;449(7159):248–51.

Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012;13(7):484–92.

Jurkowska RZ, Anspach N, Urbanke C, Jia D, Reinhardt R, Nellen W, et al. Formation of nucleoprotein filaments by mammalian DNA methyltransferase Dnmt3a in complex with regulator Dnmt3L. Nucleic Acids Res. 2008;36(21):6656–63.

Jurkowska RZ, Jurkowski TP, Jeltsch A. Structure and function of mammalian DNA methyltransferases. Chem Eur J Chem Biol. 2011a;12(2):206–22.

Jurkowska RZ, Rajavelu A, Anspach N, Urbanke C, Jankevicius G, Ragozin S, et al. Oligomerization and binding of the Dnmt3a DNA methyltransferase to parallel DNA molecules: heterochromatic localization and role of Dnmt3L. J Biol Chem. 2011b;286(27):24200–7.

Jurkowska RZ, Siddique AN, Jurkowski TP, Jeltsch A. Approaches to enzyme and substrate design of the murine Dnmt3a DNA methyltransferase. Chem Eur J Chem Biol. 2011c;12(10):1589–94.

Kareta MS, Botello ZM, Ennis JJ, Chou C, Chedin F. Reconstitution and mechanism of the stimulation of de novo methylation by human DNMT3L. J Biol Chem. 2006;281(36):25893–902.

Kim SC, Sprung R, Chen Y, Xu Y, Ball H, Pei J, et al. Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. Mol Cell. 2006;23(4):607–18.
Kim SH, Park J, Choi MC, Kim HP, Park JH, Jung Y, et al. Zinc-fingers and homeoboxes 1 (ZHX1) binds DNA methyltransferase (DNMT) 3B to enhance DNMT3B-mediated transcriptional repression. Biochem Biophys Res Commun. 2007;355(2):318–23.
Kimura H, Shiota K. Methyl-CpG-binding protein, MeCP2, is a target molecule for maintenance DNA methyltransferase, Dnmt1. J Biol Chem. 2003;278(7):4806–12.
Klimasauskas S, Kumar S, Roberts RJ, Cheng X. HhaI methyltransferase flips its target base out of the DNA helix. Cell. 1994;76(2):357–69.
Klose RJ, Bird AP. Genomic DNA methylation: the mark and its mediators. Trends Biochem Sci. 2006;31(2):89–97.
Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J. Differential chromatin marking of introns and expressed exons by H3K36me3. Nat Genet. 2009;41(3):376–81.
Larschan E, Alekseyenko AA, Gortchakov AA, Peng S, Li B, Yang P, et al. MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism. Mol Cell. 2007;28(1):121–33.
Lavoie G, Esteve PO, Laulan NB, Pradhan S, St-Pierre Y. PKC isoforms interact with and phosphorylate DNMT1. BMC Biol. 2011;9:31.
Leonhardt H, Page AW, Weier HU, Bestor TH. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell. 1992;71(5):865–73.
Li H, Rauch T, Chen ZX, Szabo PE, Riggs AD, Pfeifer GP. The histone methyltransferase SETDB1 and the DNA methyltransferase DNMT3A interact directly and localize to promoters silenced in cancer cells. J Biol Chem. 2006;281(28):19489–500.
Li JY, Pu MT, Hirasawa R, Li BZ, Huang YN, Zeng R, et al. Synergistic function of DNA methyltransferases Dnmt3a and Dnmt3b in the methylation of Oct4 and Nanog. Mol Cell Biol. 2007;27(24):8748–59.
Li BZ, Huang Z, Cui QY, Song XH, Du L, Jeltsch A, et al. Histone tails regulate DNA methylation by allosterically activating de novo methyltransferase. Cell Res. 2011;21(8):1172–81.
Lin IG, Han L, Taghva A, O'Brien LE, Hsieh CL. Murine de novo methyltransferase Dnmt3a demonstrates strand asymmetry and site preference in the methylation of DNA in vitro. Mol Cell Biol. 2002;22(3):704–23.
Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009;462(7271):315–22.
Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, et al. Global epigenomic reconfiguration during mammalian brain development. Science. 2013;341(6146):1237905.
Liu Y, Oakeley EJ, Sun L, Jost JP. Multiple domains are involved in the targeting of the mouse DNA methyltransferase to the DNA replication foci. Nucleic Acids Res. 1998;26(4):1038–45.
Liu X, Gao Q, Li P, Zhao Q, Zhang J, Li J, et al. UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. Nat Commun. 2013;4:1563.
Lukashevich OV, Cherepanova NA, Jurkovska RZ, Jeltsch A, Gromova ES. Conserved motif VIII of murine DNA methyltransferase Dnmt3a is essential for methylation activity. BMC Biochem. 2016;17(1):7.
Margot JB, Ehrenhofer-Murray AE, Leonhardt H. Interactions within the mammalian DNA methyltransferase family. BMC Mol Biol. 2003;4:7.
Margueron R, Reinberg D. Chromatin structure and the inheritance of epigenetic information. Nat Rev Genet. 2010;11(4):285–96.
Matzke MA, Mosher RA. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat Rev Genet. 2014;15(6):394–408.
Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature. 2008;454(7205):766–70.
Morselli M, Pastor WA, Montanini B, Nee K, Ferrari R, Fu K, et al. In vivo targeting of de novo DNA methylation by histone modifications in yeast and mouse. eLife. 2015;4:e06205.

Muegge K. LSH, a guardian of heterochromatin at repeat elements. Biochem Cell Biol Biochimie et biologie cellulaire. 2005;83(4):548–54.
Myant K, Stancheva I. LSH cooperates with DNA methyltransferases to repress transcription. Mol Cell Biol. 2008;28(1):215–26.
Nady N, Lemak A, Walker JR, Avvakumov GV, Kareta MS, Achour M, et al. Recognition of multivalent histone states associated with heterochromatin by UHRF1 protein. J Biol Chem. 2011;286(27):24300–11.
Neri F, Krepelova A, Incarnato D, Maldotti M, Parlato C, Galvagni F, et al. Dnmt3L antagonizes DNA methylation at bivalent promoters and favors DNA methylation at gene bodies in ESCs. Cell. 2013;155(1):121–34.
Nishiyama A, Yamaguchi L, Sharif J, Johmura Y, Kawamura T, Nakanishi K, et al. Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. Nature. 2013;502(7470):249–53.
Noh KM, Wang H, Kim HR, Wenderski W, Fang F, Li CH, et al. Engineering of a histone-recognition domain in Dnmt3a alters the epigenetic landscape and phenotypic features of mouse ESCs. Mol Cell. 2015;59(1):89–103.
O'Keefe RT, Henderson SC, Spector DL. Dynamic organization of DNA replication in mammalian cell nuclei: spatially and temporally defined replication of chromosome-specific alpha-satellite DNA sequences. J Cell Biol. 1992;116(5):1095–110.
Okano M, Xie SP, Li E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet. 1998;19(3):219–20.
Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature. 2007;448(7154):714–7.
Osipiuk J, Walsh MA, Joachimiak A. Crystal structure of MboIIA methyltransferase. Nucleic Acids Res. 2003;31(18):5440–8.
Otani J, Nankumo T, Arita K, Inamoto S, Ariyoshi M, Shirakawa M. Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. EMBO Rep. 2009;10(11):1235–41.
Peng L, Yuan Z, Ling H, Fukasawa K, Robertson K, Olashaw N, et al. SIRT1 deacetylates the DNA methyltransferase 1 (DNMT1) protein and alters its activities. Mol Cell Biol. 2011;31(23): 4720–34.
Pradhan S, Esteve PO. Allosteric activator domain of maintenance human DNA (cytosine-5) methyltransferase and its role in methylation spreading. Biochemistry. 2003;42(18):5321–32.
Pradhan M, Esteve PO, Chin HG, Samaranayke M, Kim GD, Pradhan S. CXXC domain of human DNMT1 is essential for enzymatic activity. Biochemistry. 2008;47(38):10000–9.
Purdy MM, Holz-Schietinger C, Reich NO. Identification of a second DNA binding site in human DNA methyltransferase 3A by substrate inhibition and domain deletion. Arch Biochem Biophys. 2010;498(1):13–22.
Qin S, Min J. Structure and function of the nucleosome-binding PWWP domain. Trends Biochem Sci. 2014;39(11):536–47.
Qin W, Leonhardt H, Spada F. Usp7 and Uhrf1 control ubiquitination and stability of the maintenance DNA methyltransferase Dnmt1. J Cell Biochem. 2011;112(2):439–44.
Qin W, Wolf P, Liu N, Link S, Smets M, La Mastra F, et al. DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination. Cell Res. 2015;25(8):911–29.
Qiu C, Sawada K, Zhang X, Cheng X. The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. Nat Struct Biol. 2002;9(3):217–24.
Quenneville S, Verde G, Corsinotti A, Kapopoulou A, Jakobsson J, Offner S, et al. In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. Mol Cell. 2011;44(3):361–72.
Rajakumara E, Wang Z, Ma H, Hu L, Chen H, Lin Y, et al. PHD finger recognition of unmodified histone H3R2 links UHRF1 to regulation of euchromatic gene expression. Mol Cell. 2011;43(2):275–84.
Rajavelu A, Jurkowska RZ, Fritz J, Jeltsch A. Function and disruption of DNA methyltransferase 3a cooperative DNA binding and nucleoprotein filament formation. Nucleic Acids Res. 2012;40(2):569–80.

Ramsahoye BH, Biniszkiewicz D, Lyko F, Clark V, Bird AP, Jaenisch R. Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc Natl Acad Sci U S A. 2000;97(10):5237–42.

Reither S, Li F, Gowher H, Jeltsch A. Catalytic mechanism of DNA-(cytosine-C5)-methyltransferases revisited: covalent intermediate formation is not essential for methyl group transfer by the murine Dnmt3a enzyme. J Mol Biol. 2003;329(4):675–84.

Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. Annu Rev Biochem. 2012; 81:145–66.

Robertson KD, Uzvolgyi E, Liang G, Talmadge C, Sumegi J, Gonzales FA, et al. The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal tissues and overexpression in tumors. Nucleic Acids Res. 1999;27(11):2291–8.

Rondelet G, Dal Maso T, Willems L, Wouters J. Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B. J Struct Biol. 2016;194(3):357–67.

Rothbart SB, Krajewski K, Nady N, Tempel W, Xue S, Badeaux AI, et al. Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. Nat Struct Mol Biol. 2012;19(11):1155–60.

Rothbart SB, Dickson BM, Ong MS, Krajewski K, Houliston S, Kireev DB, et al. Multivalent histone engagement by the linked tandem Tudor and PHD domains of UHRF1 is required for the epigenetic inheritance of DNA methylation. Genes Dev. 2013;27(11):1288–98.

Scavetta RD, Thomas CB, Walsh MA, Szegedi S, Joachimiak A, Gumport RI, et al. Structure of RsrI methyltransferase, a member of the N6-adenine beta class of DNA methyltransferases. Nucleic Acids Res. 2000;28(20):3950–61.

Schmitz KM, Mayer C, Postepska A, Grummt I. Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. Genes Dev. 2010;24(20):2264–9.

Schultz MD, He Y, Whitaker JW, Hariharan M, Mukamel EA, Leung D, et al. Human body epigenome maps reveal noncanonical DNA methylation variation. Nature. 2015;523(7559):212–6.

Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, Endo TA, et al. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature. 2007; 450(7171):908–12.

Sharma S, De Carvalho DD, Jeong S, Jones PA, Liang G. Nucleosomes containing methylated DNA stabilize DNA methyltransferases 3A/3B and ensure faithful epigenetic inheritance. PLoS Genet. 2011;7(2):e1001286.

Smallwood SA, Tomizawa S, Krueger F, Ruf N, Carli N, Segonds-Pichon A, et al. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat Genet. 2011;43(8): 811–4.

Song J, Rechkoblit O, Bestor TH, Patel DJ. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. Science. 2011;331(6020):1036–40.

Song J, Teplova M, Ishibe-Murakami S, Patel DJ. Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. Science. 2012;335(6069):709–12.

Suetake I, Shinozaki F, Miyagawa J, Takeshima H, Tajima S. DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. J Biol Chem. 2004;279(26):27816–23.

Suetake I, Hayata D, Tajima S. The amino-terminus of mouse DNA methyltransferase 1 forms an independent domain and binds to DNA with the sequence involving PCNA binding motif. J Biochem. 2006;140(6):763–76.

Suetake I, Mishima Y, Kimura H, Lee YH, Goto Y, Takeshima H, et al. Characterization of DNA-binding activity in the N-terminal domain of the DNA methyltransferase Dnmt3a. Biochem J. 2011;437(1):141–8.

Sugiyama Y, Hatano N, Sueyoshi N, Suetake I, Tajima S, Kinoshita E, et al. The DNA-binding activity of mouse DNA methyltransferase 1 is regulated by phosphorylation with casein kinase 1delta/epsilon. Biochem J. 2010;427(3):489–97.

Suva ML, Riggi N, Bernstein BE. Epigenetic reprogramming in cancer. Science. 2013;339(6127): 1567–70.

Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet. 2008;9(6):465–76.
Suzuki M, Yamada T, Kihara-Negishi F, Sakurai T, Hara E, Tenen DG, et al. Site-specific DNA methylation by a complex of PU.1 and Dnmt3a/b. Oncogene. 2006;25(17):2477–88.
Svedruzic ZM, Reich NO. DNA cytosine C5 methyltransferase Dnmt1: catalysis-dependent release of allosteric inhibition. Biochemistry. 2005;44(27):9472–85.
Syeda F, Fagan RL, Wean M, Avvakumov GV, Walker JR, Xue S, et al. The replication focus targeting sequence (RFTS) domain is a DNA-competitive inhibitor of Dnmt1. J Biol Chem. 2011;286(17):15344–51.
Takeshima H, Suetake I, Tajima S. Mouse Dnmt3a preferentially methylates linker DNA and is inhibited by histone H1. J Mol Biol. 2008;383(4):810–21.
Takeshita K, Suetake I, Yamashita E, Suga M, Narita H, Nakagawa A, et al. Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). Proc Natl Acad Sci U S A. 2011;108(22):9055–9.
Vakoc CR, Sachdeva MM, Wang H, Blobel GA. Profile of histone lysine methylation across transcribed mammalian chromatin. Mol Cell Biol. 2006;26(24):9185–95.
van Nuland R, van Schaik FM, Simonis M, van Heesch S, Cuppen E, Boelens R, et al. Nucleosomal DNA binding drives the recognition of H3K36-methylated nucleosomes by the PSIP1-PWWP domain. Epigenetics Chromatin. 2013;6(1):12.
Varley KE, Gertz J, Bowling KM, Parker SL, Reddy TE, Pauli-Behn F, et al. Dynamic DNA methylation across diverse human cell lines and tissues. Genome Res. 2013;23(3):555–67.
Vilkaitis G, Suetake I, Klimasauskas S, Tajima S. Processive methylation of hemimethylated CpG sites by mouse Dnmt1 DNA methyltransferase. J Biol Chem. 2005;280(1):64–72.
Vire E, Brenner C, Deplus R, Blanchon L, Fraga M, Didelot C, et al. The Polycomb group protein EZH2 directly controls DNA methylation. Nature. 2006;439(7078):871–4.
Wang J, Hevi S, Kurash JK, Lei H, Gay F, Bajko J, et al. The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. Nat Genet. 2009;41(1):125–9.
Wang C, Shen J, Yang Z, Chen P, Zhao B, Hu W, et al. Structural basis for site-specific reading of unmodified R2 of histone H3 tail by UHRF1 PHD finger. Cell Res. 2011;21(9):1379–82.
Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet. 2007;39(4):457–66.
Zhang Y, Rohde C, Tierling S, Jurkowski TP, Bock C, Santacruz D, et al. DNA methylation analysis of chromosome 21 gene promoters at single base pair and single allele resolution. PLoS Genet. 2009;5(3):e1000438.
Zhang Y, Jurkowska R, Soeroes S, Rajavelu A, Dhayalan A, Bock I, et al. Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. Nucleic Acids Res. 2010;38(13):4246–53.
Zhang G, Esteve PO, Chin HG, Terragni J, Dai N, Correa Jr IR, et al. Small RNA-mediated DNA (cytosine-5) methyltransferase 1 inhibition leads to aberrant DNA methylation. Nucleic Acids Res. 2015a;43(12):6112–24.
Zhang ZM, Liu S, Lin K, Luo Y, Perry JJ, Wang Y, et al. Crystal structure of human DNA methyltransferase 1. J Mol Biol. 2015b;427(15):2520–31.
Zhou Q, Agoston AT, Atadja P, Nelson WG, Davidson NE. Inhibition of histone deacetylases promotes ubiquitin-dependent proteasomal degradation of DNA methyltransferase 1 in human breast cancer cells. Mole Cancer Res MCR. 2008;6(5):873–83.
Zhu H, Geiman TM, Xi S, Jiang Q, Schmidtmann A, Chen T, et al. Lsh is involved in de novo methylation of DNA. EMBO J. 2006;25(2):335–45.
Ziller MJ, Muller F, Liao J, Zhang Y, Gu H, Bock C, et al. Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. PLoS Genet. 2011;7(12):e1002389.

Genetic Studies on Mammalian DNA Methyltransferases

Jiameng Dan and Taiping Chen

Abstract

Cytosine methylation at the C5-position, generating 5-methylcytosine (5mC), is a DNA modification found in many eukaryotic organisms, including fungi, plants, invertebrates, and vertebrates, albeit its levels vary greatly in different organisms. In mammals, cytosine methylation occurs predominantly in the context of CpG dinucleotides, with the majority (60–80 %) of CpG sites in their genomes being methylated. DNA methylation plays crucial roles in the regulation of chromatin structure and gene expression and is essential for mammalian development. Aberrant changes in DNA methylation levels and patterns are associated with various human diseases, including cancer and developmental disorders. DNA methylation is mediated by three active DNA methyltransferases (Dnmts), namely, Dnmt1, Dnmt3a, and Dnmt3b, in mammals. Over the last two decades, genetic manipulations of these enzymes, as well as their regulators, in mice have greatly contributed to our understanding of the biological functions of DNA methylation in mammals. In this chapter, we discuss genetic studies on

J. Dan
Department of Epigenetics and Molecular Carcinogenesis, The University of Texas MD Anderson Cancer Center, 1808 Park Road 1C, Smithville, TX 78957, USA

Center for Cancer Epigenetics, The University of Texas MD Anderson Cancer Center, 1808 Park Road 1C, Smithville, TX 78957, USA

T. Chen (✉)
Department of Epigenetics and Molecular Carcinogenesis, The University of Texas MD Anderson Cancer Center, 1808 Park Road 1C, Smithville, TX 78957, USA

Center for Cancer Epigenetics, The University of Texas MD Anderson Cancer Center, 1808 Park Road 1C, Smithville, TX 78957, USA

Graduate School of Biomedical Sciences at Houston, Houston, TX 77030, USA
e-mail: tchen2@mdanderson.org

© Springer International Publishing Switzerland 2016

A. Jeltsch, R.Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*, Advances in Experimental Medicine and Biology 945,
DOI 10.1007/978-3-319-43624-1_6

mammalian Dnmts, focusing on their roles in embryogenesis, cellular differentiation, genomic imprinting, and X-chromosome inactivation.

Abbreviations

5caC	5-Carboxylcytosine
5fC	5-Formylcytosine
5hmC	5-Hydroxymethylcytosine
5mC	5-Methylcytosine
ADCA-DN	Autosomal dominant cerebellar ataxia deafness and narcolepsy
ADD	ATRX-Dnmt3-Dnmt3L
AML	Acute myeloid leukemia
BAH	Bromo-adjacent homology
DKO	Double knockout
DMR	Differentially methylated region
DNMT	DNA methyltransferase
ES	Embryonic stem
EST	Expressed sequence tag
HP1	Heterochromatin protein 1
HSAN IE	Hereditary sensory and autonomic neuropathy with dementia and hearing loss type IE
ICF	Immunodeficiency centromeric instability and facial anomalies
ICM	Inner cell mass
ICR	Imprinting control region
KAP1	KRAB-associated protein 1
KRAB	*Krüppel*-associated box
lncRNA	Long non-coding RNA
MBD3	Methyl CpG-binding domain protein-3
MEF	Mouse embryonic fibroblast
MTA2	Metastasis tumor antigen 2
NLS	Nuclear localization signal
NuRD	Nuclear remodeling and histone deacetylation
PBD	PCNA-binding domain
PCNA	Proliferating cell nuclear antigen
PGC	Primordial germ cell
PHD	Plant homeodomain
PRC2	Polycomb repressive complex 2
PWWP	Proline-tryptophan-tryptophan-proline
RFTS	Replication foci-targeting sequence
RING	Really Interesting New Gene
SRA	SET- and RING-associated
TDG	Thymine DNA glycosylase
TTD	Tandem tudor domain
UBL	Ubiquitin-like

Uhrf1	Ubiquitin-like with PHD and RING finger domains 1
Xa	Active X chromosome
XCI	X-chromosome inactivation
Xi	Inactive X chromosome
Xic	X-inactivation center
Xist	X-inactive-specific transcript
Xm	Maternal X chromosome
Xp	Paternal X chromosome

1 Distinct Roles of Dnmt1 and Dnmt3 Families in DNA Methylation

In 1975, long before the identification of any mammalian DNA methyltransferase, Holliday and Pugh and Riggs independently proposed a theory that DNA methylation could serve as a heritable epigenetic mark for cellular memory. Recognizing that the CpG dinucleotide is self-complementary, they postulated that methylated and unmethylated CpG sites could be copied when cells divide so that DNA methylation patterns would be replicated semiconservatively like the base sequence of DNA itself (Holliday and Pugh 1975; Riggs 1975). A prediction of the theory was the existence of at least two DNA methyltransferase activities: *de novo* methyltransferase(s) would methylate unmodified DNA and establish DNA methylation patterns, and maintenance methyltransferase(s) would recognize hemimethylated sites and "copy" the methylation patterns from the parental strand onto the daughter strand at each round of DNA replication.

1.1 Dnmt1: The Major Maintenance Methyltransferase

The first mammalian DNA methyltransferase gene, *Dnmt1*, was cloned from murine cells (Bestor et al. 1988). The *Dnmt1* locus has several transcription start sites and produces two major protein products (Mertineit et al. 1998; Rouleau et al. 1992). Transcription initiation within a somatic cell-specific exon (exon 1 s) results in the Dnmt1s isoform (generally referred to as Dnmt1) which consists of 1620 amino acids. Transcription initiation within an oocyte-specific exon (exon 1o) produces a transcript that utilizes a downstream AUG as the translation initiation codon. As a result, the oocyte-specific isoform, Dnmt1o, lacks the N-terminal 118 amino acids of Dnmt1s (Mertineit et al. 1998). Although Dnmt1o is more stable than Dnmt1s, genetic evidence suggests no functional difference between these isoforms (Ding and Chaillet 2002). Human DNMT1, consisting of 1616 amino acids, is nearly 80 % identical to the mouse Dnmt1 at the amino acid level.

Dnmt1 contains a C-terminal catalytic domain containing characteristic amino acid sequence motifs that are homologous to bacterial DNA methyltransferases and an N-terminal regulatory region that is not present in bacterial enzymes (Bestor

et al. 1988). The N-terminal regulatory region contains several functional domains, including a proliferating cell nuclear antigen (PCNA)-binding domain (PBD) responsible for the interaction with the DNA replication machinery, a nuclear localization signal (NLS), a replication foci-targeting sequence (RFTS) that mediates the association with late replicating heterochromatin, a zinc-finger CXXC domain that recognizes unmethylated CpG-containing DNA, and a pair of bromo-adjacent homology (BAH) domains (Fig. 1a). Recent structural data revealed that the RFTS domain binds to the catalytic domain and blocks the catalytic center, suggesting an autoinhibitory role in the regulation of enzymatic activity (Takeshita et al. 2011).

In vitro biochemical assays revealed that, although Dnmt1 is capable of methylating both unmethylated and hemimethylated CpG dinucleotides, its activity toward hemimethylated substrates is far more efficient (Pradhan et al. 1999). Dnmt1 is ubiquitously expressed through development, with high levels in proliferating cells. Dnmt1 associates with the DNA replication machinery at S phase and with heterochromatin at late S and G2 phases (Chuang et al. 1997; Easwaran et al. 2004; Leonhardt et al. 1992; Schneider et al. 2013), suggesting that Dnmt1-mediated methylation is coupled to DNA replication. These findings supported the notion that Dnmt1 functions as a maintenance enzyme (Fig. 1b). However, because Dnmt1, the only known DNA methyltransferase at the time, also has *de novo* methylation activity *in vitro*, it was initially debated whether *de novo* methylation and maintenance methylation are carried out by Dnmt1 alone or by two or more distinct enzymes (Bestor 1992).

Genetic studies in mouse models and murine cells helped settling the debate. Several *Dnmt1* mutant alleles were generated by gene targeting. The $Dnmt1^{n}$ allele (n stands for N-terminal disruption) was reported in 1992 (Li et al. 1992). This allele, in which a genomic region coding 60 amino acids near the N-terminal end was replaced by a neomycin resistance cassette, is a partial loss-of-function (hypomorphic) mutation. $Dnmt1^{n/n}$ embryos have a ~70 % reduction in global DNA methylation and show mid-gestation lethality (Li et al. 1992). Subsequently, the $Dnmt1^{s}$ allele (s stands for *Sal*I site) was reported, which had a neomycin resistance cassette inserted into a *Sal*I site in exon 17, disrupting the RFTS (Li et al. 1993). The $Dnmt1^{s}$ allele is functionally more severe than the $Dnmt1^{n}$ allele, as $Dnmt1^{s/s}$ embryos show lower levels of DNA methylation and earlier lethality (Lei et al. 1996). However, it was unclear whether the $Dnmt1^{s}$ allele was a null mutation, because the C-terminal catalytic domain was intact. To completely inactivate *Dnmt1*, Lei et al. generated the $Dnmt1^{c}$ allele (c stands for C-terminal disruption) by disrupting the catalytic domain, including the highly conserved PC and ENV motifs that are essential for enzymatic activity (Lei et al. 1996). The development of $Dnmt1^{c/c}$ embryos is arrested prior to the 8-somite stage, significantly earlier than the developmental phenotype of $Dnmt1^{n/n}$ embryos, while the viability and proliferation of *Dnmt1* null embryonic stem (ES) cells are not affected (Lei et al. 1996). Inactivation of Dnmt1 by mutating the cysteine (C1229) residue at the catalytic center (PC motif) results in similar developmental defects (Takebayashi et al. 2007), suggesting that the phenotype is largely due to the loss of catalytic activity. DNA methylation analyses revealed that *Dnmt1* null embryos and ES cells contain low but stable levels of

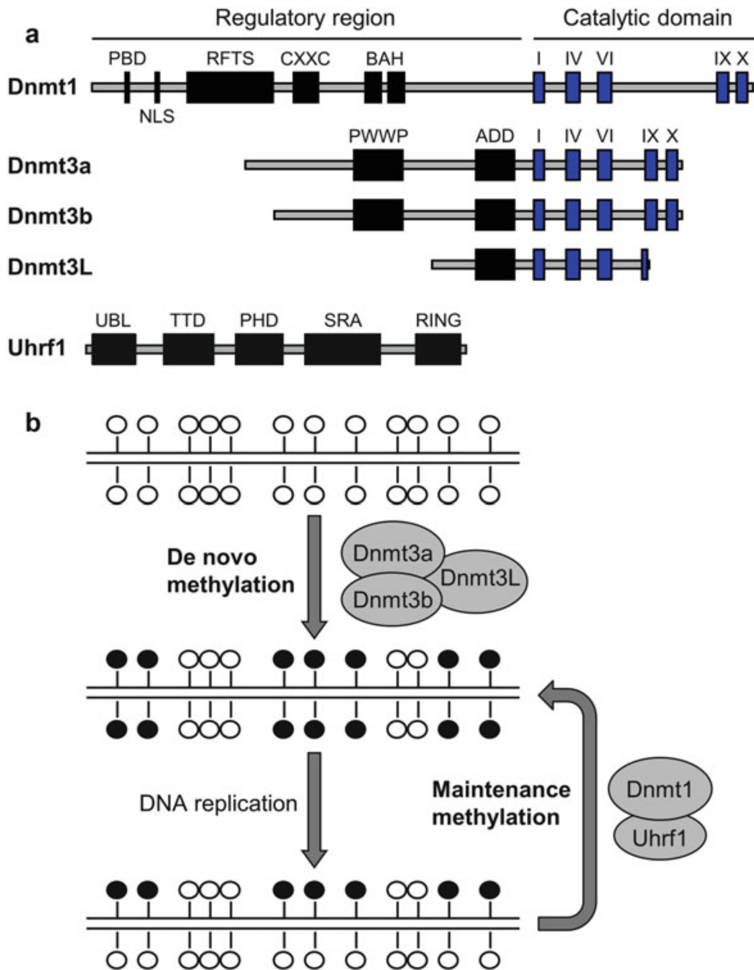


Fig. 1 DNA methyltransferases and major regulatory proteins involved in DNA methylation. (**a**) Schematic diagrams of Dnmt1, Dnmt3a, Dnmt3b, Dnmt3L, and Uhrf1. The C-terminal catalytic domains of the Dnmt1 and Dnmt3 families are conserved (the highly conserved signature motifs I, IV, VI, IX, and X are shown), but their N-terminal regulatory regions are distinct. Functional domains of the proteins are indicated. *PBD* PCNA-binding domain, *NLS* nuclear localization signal, *RFTS* replication foci-targeting sequence, *CXXC* a cysteine-rich domain implicated in binding CpG-containing DNA sequences, *BAH* bromo-adjacent homology domain, *PWWP* proline-tryptophan-tryptophan-proline domain, *ADD* ATRX-Dnmt3-Dnmt3L domain, and *UBL* ubiquitin-like domain; *TTD* tandem tudor domain, *PHD* plant homeodomain, *SRA* SET- and RING-associated domain, and *RING* Really Interesting New Gene domain. (**b**) *De novo* and maintenance methyltransferase activities. The *de novo* methyltransferases Dnmt3a and Dnmt3b, in complex with their accessory factor Dnmt3L, methylate unmodified DNA and establish methylation patterns. At each round of DNA replication, the maintenance methyltransferase Dnmt1, aided by its accessory factor Uhrf1, “copies” the methylation pattern from the parental strand onto the daughter strand. Open circles represent unmethylated CpG dinucleotides, and filled circles represent methylated CpG dinucleotides

5-methylcytosine (5mC) and methyltransferase activity. Moreover, the *de novo* methylation activity is not impaired by Dnmt1 loss, as integrated provirus DNA in MoMuLV-infected *Dnmt1* null ES cells becomes methylated at a similar rate as in wild-type ES cells (Lei et al. 1996). Taken together, these studies provided compelling evidence for the existence of one or more DNA methyltransferases that are important for *de novo* methylation.

1.2 Dnmt2/Trdmt1: A tRNA Methyltransferase

Results from genetic studies of *Dnmt1* prompted the search for more DNA methyltransferase genes. In 1998, several groups reported the identification of a second putative DNA methyltransferase gene, named *Dnmt2*, which encodes a protein of 391 amino acids in human or 415 amino acids in mouse (Okano et al. 1998b; Van den Wyngaert et al. 1998; Yoder and Bestor 1998). Despite the presence of all the conserved motifs shared by known prokaryotic and eukaryotic DNA cytosine methyltransferases, Dnmt2 has no detectable DNA methyltransferase activity in standard *in vitro* assays. Furthermore, inactivation of *Dnmt2* in mouse ES cells by gene targeting has no effect on preexisting genomic methylation patterns or on the ability to methylate newly integrated retrovirus DNA *de novo* (Okano et al. 1998b). Indeed, a subsequent study demonstrated that Dnmt2 is a tRNA methyltransferase, specific for cytosine 38 in the anticodon loop of aspartic acid tRNA, and has been renamed tRNA aspartic acid (D) methyltransferase 1 (Trdmt1) (Goll et al. 2006).

1.3 Dnmt3a and Dnmt3b: The *De Novo* Methyltransferases

By searching an expressed sequence tag (EST) database using full-length bacterial type II cytosine-C5 methyltransferase sequences as queries, Okano *et al*. identified two additional homologous genes, *Dnmt3a* and *Dnmt3b*, in both mouse and human. Their protein products contain the highly conserved DNA methyltransferase motifs in their C-terminal regions, but their N-terminal regulatory regions are unrelated to that of Dnmt1 (Okano et al. 1998a). The N-terminal regions of Dnmt3a and Dnmt3b contain a variable region and two conserved domains, the proline-tryptophan-tryptophan-proline (PWWP) domain and the ATRX-Dnmt3-Dnmt3L (ADD) domain (Fig. 1a). Both domains are implicated in chromatin binding. The PWWP domain is required for heterochromatin localization and mediates H3K36me3 binding (Baubec et al. 2015; Chen et al. 2004; Dhayalan et al. 2010). The ADD domain interacts with the N-terminal tail of histone H3, and the interaction is disrupted by various posttranslational modifications of H3, including di- and trimethylation of K4, acetylation of K4, and phosphorylation of T3, S10, or T11 (Otani et al. 2009; Noh et al. 2015; Zhang et al. 2010).

Dnmt3a produces two major isoforms, Dnmt3a and Dnmt3a2, driven by different promoters (Chen et al. 2002). The full-length Dnmt3a protein, consisting of

908 amino acids in mouse and 912 amino acids in human, is expressed ubiquitously at relatively low levels. The *Dnmt3a2* transcript is initiated in intron 6 of the *Dnmt3a* gene and encodes a protein that lacks the N-terminal 219 (in mouse) or 223 (in human) amino acids of Dnmt3a. Dnmt3a2, which is catalytically active, is the predominant form in mouse ES cells, early embryos, and developing germ cells, as well as human embryonal carcinoma cells, and is also detectable in spleen and thymus (Chen et al. 2002). The *Dnmt3b* gene produces multiple alternatively spliced isoforms, some of which encode catalytically inactive protein products. The longest isoform, Dnmt3b1, consists of 859 amino acids in mouse and 853 amino acids in human, respectively. Both active and inactive Dnmt3b isoforms appear to co-express in most, if not all, cell types. For example, Dnmt3b1, an active form, and Dnmt3b6, an inactive form, are the predominant forms in mouse ES cells, whereas Dnmt3b2, an active form, and Dnmt3b3, an inactive form, are expressed at low levels in many somatic cells (Chen et al. 2002). There is evidence that catalytically inactive Dnmt3b protein products may play regulatory roles in DNA methylation. For example, overexpression of human DNMT3B7, a truncated isoform frequently found in cancer cells, leads to higher levels of total genomic methylation and altered gene expression in both transgenic mice and human cancer cells (Ostler et al. 2012; Shah et al. 2010).

Several lines of evidence suggest the involvement of Dnmt3a and Dnmt3b in *de novo* DNA methylation (Fig. 1b). First, Dnmt3a and Dnmt3b are highly expressed in early embryos (and ES cells) and developing germ cells, where an active *de novo* methylation takes place, but are downregulated in somatic tissues and when ES cells are induced to differentiate (Okano et al. 1998a). Second, recombinant Dnmt3a and Dnmt3b proteins methylate unmethylated and hemimethylated DNA with equal efficiency (Okano et al. 1998a). Genetic studies provided definitive evidence that Dnmt3a and Dnmt3b were the long-sought *de novo* methyltransferases. Inactivation of both *Dnmt3a* and *Dnmt3b* by gene targeting blocks *de novo* methylation in ES cells and early embryos but has no effect on maintenance of imprinted methylation patterns (Okano et al. 1999). Dnmt3a deficiency also leads to failure to establish DNA methylation imprints in developing germ cells (Kaneda et al. 2004).

It is worth noting that the *de novo* DNA methyltransferase activity of Dnmt3a and Dnmt3b is not only essential for the establishment of new DNA methylation patterns but also important for the faithful maintenance of these patterns. In culture, *Dnmt3a*/*3b* double knockout (DKO) ES cells exhibit gradual loss of global DNA methylation and, after multiple passages, show severe hypomethylation (Chen et al. 2003), suggesting that Dnmt1 and Dnmt3 enzymes have distinct and nonredundant functions but act cooperatively in the maintenance of global DNA methylation. Based on the kinetics of DNA methylation loss, it was proposed that Dnmt1 is the major maintenance methyltransferase that, upon DNA replication, methylates hemimethylated CpG sites with high efficiency but not absolute fidelity, whereas Dnmt3a and Dnmt3b, as *de novo* methyltransferases, act as "proofreading" enzymes that methylate the hemimethylated CpG sites missed by Dnmt1 (Chen et al. 2003).

1.4 Dnmt3L: A Regulator of *De Novo* Methylation

A third member of the Dnmt3 family, *Dnmt3*-like (*Dnmt3L*), was originally isolated by database analysis of the human genome sequence (Aapola et al. 2000). Its murine homolog was subsequently identified (Aapola et al. 2001; Hata et al. 2002). The human and mouse Dnmt3L proteins consist of 387 and 421 amino acids, respectively. Dnmt3L contains an ADD domain, but lacks a PWWP domain, in the N-terminal region. Its C-terminal region is highly related to the catalytic domains of Dnmt3a and Dnmt3b, but lacks some motifs essential for enzymatic activity, including the PC dipeptide at the active site and the sequence motif involved in binding of the methyl donor *S*-adenosyl-L-methionine (Aapola et al. 2000, 2001; Hata et al. 2002) (Fig. 1a). Therefore, Dnmt3L has no methyltransferase activity. However, Dnmt3L has been shown to interact with Dnmt3a and Dnmt3b, stimulate their enzymatic activities, and target them to chromatin (Hata et al. 2002; Jia et al. 2007; Ooi et al. 2007; Suetake et al. 2004). The expression pattern of Dnmt3L during development is also strikingly similar to that of Dnmt3a and Dnmt3b, including high expression in developing germ cells, early embryos, and ES cells (Hata et al. 2002). These findings indicate that Dnmt3L may regulate Dnmt3a/3b functions (Fig. 1b). Genetic studies indeed demonstrate that Dnmt3L is an essential accessory factor of Dnmt3a in the germ line. *Dnmt3L* homozygous null mice are viable and grossly normal, but both male and female mice are infertile (Bourc'his et al. 2001; Hata et al. 2002). Male mice show activation of retrotransposons in spermatogonia and spermatocytes, due to failure to establish methylation at these elements, and are azoospermic (Bourc'his and Bestor 2004). Female mice fail to establish maternal methylation imprints in oocytes, and, as a result, embryos derived from these oocytes cannot survive beyond mid-gestation (Bourc'his et al. 2001; Hata et al. 2002). The phenotype is indistinguishable from that of mice with conditional *Dnmt3a* deletion in germ cells (Kaneda et al. 2004). Recently, Dnmt3L was shown to antagonize DNA methylation at H3K4me3/K27me3 bivalent promoters, which are often associated with developmental genes, and favor DNA methylation at gene bodies in ES cells. It was suggested that Dnmt3L, via its ADD domain, interacts with Polycomb repressive complex 2 (PRC2) in competition with Dnmt3a and Dnmt3b to maintain low methylation levels at regions with H3K27me3, thus maintaining hypomethylation at promoters of bivalent developmental genes (Neri et al. 2013). The physiological relevance of this finding remains to be determined, given that zygotic Dnmt3L is not required for embryonic development and postnatal survival (Bourc'his et al. 2001; Hata et al. 2002).

1.5 Uhrf1: A Regulator of Maintenance Methylation

Besides Dnmts, a number of DNA methylation regulators have been identified, including the multi-domain protein Uhrf1 (ubiquitin-like with PHD and RING finger domains 1), also known as NP95 (mouse) and ICBP90 (human) (Fig. 1a). Genetic studies demonstrated an essential role for Uhrf1 in maintaining DNA

methylation (Fig. 1b). Uhrf1 deficiency leads to embryonic lethality and global DNA hypomethylation (Bostick et al. 2007; Muto et al. 2002; Sharif et al. 2007), resembling the phenotype of Dnmt1 deficiency. Cellular and biochemical evidence suggested functional interactions between Uhrf1 and Dnmt1. Uhrf1 co-localizes with Dnmt1 at DNA replication foci and heterochromatin, and Dnmt1 fails to enrich at these regions in the absence of Uhrf1 (Bostick et al. 2007; Liu et al. 2013; Sharif et al. 2007). These findings suggest that Uhrf1 is a key accessory factor for directing Dnmt1 to hemimethylated CpG sites. However, it remains somewhat controversial as to whether Uhrf1 directly recruits Dnmt1 or indirectly controls Dnmt1 localization by affecting chromatin structure. Uhrf1 harbors five known functional domains: a ubiquitin-like domain (UBL) at the N-terminus, followed by a tandem tudor domain (TTD), a plant homeodomain (PHD), a SET- and RING-associated (SRA) domain, and a Really Interesting New Gene (RING) domain (Fig. 1a). All the domains, with the exception of UBL, have been shown to be important for Dnmt1 subnuclear localization and maintenance of DNA methylation. Biochemical and structural evidence revealed that the SRA domain preferentially binds hemimethylated DNA and is thought to play an important role in loading Dnmt1 onto newly synthesized DNA substrates (Arita et al. 2008; Avvakumov et al. 2008; Bostick et al. 2007; Hashimoto et al. 2008; Sharif et al. 2007). The association of Uhrf1 with heterochromatin is also mediated by TTD, which contains an aromatic cage for binding of the heterochromatic H3K9me3 mark. The PHD acts in combination with TTD to read the H3K9me3 mark and, additionally, interacts with histone H3 tails with unmethylated R2 (H3R2me0) (Cheng et al. 2013; Liu et al. 2013; Rothbart et al. 2012, 2013; Rottach et al. 2010). Recent studies suggested that Uhrf1, via the E3 ligase activity of its RING domain, mediates ubiquitylation of H3K23 and H3K18, creating binding sites for Dnmt1 (Nishiyama et al. 2013; Qin et al. 2015). It is worth noting that Uhrf1 has also been shown to control Dnmt1 ubiquitylation and stability (Du et al. 2010; Qin et al. 2011). Indeed, a recent study revealed that Uhrf1 overexpression results in DNA hypomethylation, due to destabilization and delocalization of Dnmt1, which led the authors to propose that Uhrf1 overexpression, which is frequently observed in cancer cells, is a mechanism underlying global DNA hypomethylation in cancer (Mudbhary et al. 2014).

2 Dnmts in Embryonic Development and Cellular Differentiation

2.1 Dynamic Changes of DNA Methylation During Early Embryogenesis

DNA methylation is relatively stable in somatic tissues but exhibits dynamic changes in early embryos. During preimplantation development, both the maternal and paternal genomes undergo global DNA demethylation, albeit the mechanisms involved are distinct. The maternal genome is demethylated mainly through DNA replication-dependent passive dilution because of deficient maintenance

methylation, presumably due to the exclusion of Dnmt1 from the nucleus (Hirasawa et al. 2008; Howell et al. 2001). In contrast, demethylation of the paternal genome involves both active and passive mechanisms. Shortly after fertilization and before the first cell division, the 5mC dioxygenase Tet3 converts the majority of 5mC in the male pronucleus to 5-hydroxymethylcytosine (5hmC) (Gu et al. 2011; Wossidlo et al. 2011). 5hmC can be further oxidized to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), which can be excised by thymine DNA glycosylase (TDG) and replaced by unmodified cytosine (He et al. 2011; Ito et al. 2011). 5hmC, 5fC, and 5caC can also be passively diluted during cleavage divisions (Inoue et al. 2011; Inoue and Zhang 2011). As a result of these processes, DNA methylation marks inherited from gametes are largely erased by the blastocyst stage, with the exception of imprinting control regions (ICRs) and some retroelements, which resist this wave of global demethylation (Borgel et al. 2010; Smith et al. 2012). Around the time of implantation, *de novo* methylation takes place when the inner cell mass (ICM) starts to differentiate to form the embryonic ectoderm. Lineage-specific DNA methylation patterns are then stably maintained.

2.2 Embryonic and Adult Phenotypes of Dnmt Mutant Mice

Most of our knowledge about the significance of DNA methylation in mammalian development came from genetic manipulations of *Dnmt* genes in mice. Results from characterization of *Dnmt* mutant mice demonstrated that the establishment of embryonic methylation patterns requires both *de novo* and maintenance Dnmts and that maintaining genomic methylation above a threshold level is essential for embryonic development (Lei et al. 1996; Li et al. 1992; Okano et al. 1999). Complete inactivation of Dnmt1 results in the arrest of embryonic development between presomite and 8-somite stage around E9.5 (Lei et al. 1996). DNA methylation analysis showed that embryos deficient for Dnmt1 undergo dramatic decreases in global DNA methylation (Lei et al. 1996; Li et al. 1992), in agreement with its role as the major maintenance Dnmt. Disruption of *Dnmt3b* also leads to embryonic lethality after E12.5, with multiple defects, including growth impairment and rostral neural tube defects (Okano et al. 1999). In contrast, *Dnmt3a*-deficient mice develop to term and appear to be normal at birth but become runted and die at about 4 weeks (Okano et al. 1999). Consistent with the developmental phenotypes, DNA methylation analysis of E9.5 embryos revealed that germ line-specific genes, pluripotency genes, hematopoietic genes, and eye genes are severely hypomethylated in the absence of Dnmt3b but not of Dnmt3a (Borgel et al. 2010). This suggested that Dnmt3b is the main enzyme responsible for *de novo* methylation during embryogenesis. Dnmt3b shows a dynamic expression change during pre- and early postimplantation development, with preferential expression in the trophectoderm at the mid-blastocyst stage and subsequent transition of expression in the embryonic lineage (Hirasawa and Sasaki 2009). Notably, DNA methylation at certain genes such as *Brdt*, *Dpep3*, *Cytip*, and *Crygd* is only partially reduced in *Dnmt3b*$^{-/-}$ embryos (Borgel et al. 2010), suggesting that Dnmt3a cooperates with Dnmt3b to methylate some loci. Indeed,

Dnmt3a/3b DKO embryos exhibit more severe defects than *Dnmt3b*$^{-/-}$ embryos. Specifically, DKO embryos show smaller size, lack somites, do not undergo embryonic turning, and die before E11.5, indicating that their growth and morphogenesis are arrested shortly after gastrulation (Okano et al. 1999).

Conditional knockout (KO) studies have also demonstrated that Dnmts and DNA methylation are essential in various organs and tissues. For example, disruption of both *Dnmt1* and *Dnmt3a* in forebrain excitatory neurons leads to abnormal synaptic plasticity and deficits in learning and memory (Feng et al. 2010). Conditional deletion of *Dnmt1* at sequential stages of T cell development has also revealed a critical role for DNA methylation in T cell development, function, and survival. Specifically, deletion of *Dnmt1* in early double-negative thymocytes leads to an impaired survival of TCRαβ(+) cells and the generation of atypical CD8(+) TCRγδ(+) cells and deletion of Dnmt1 in double-positive thymocytes impairs activation-induced proliferation but differentially enhanced cytokine mRNA expression by naive peripheral T cells (Lee et al. 2001).

2.3 Cellular Defects of Dnmt Mutations

The mechanisms underlying the developmental defects observed in *Dnmt* mutant mice are not fully understood. Dnmt1, Dnmt3a, and Dnmt3b are all highly expressed in pluripotent ES cells, but disruption of these genes individually, both *Dnmt3a* and *Dnmt3b*, or even all three *Dnmts*, has no deleterious effects on mouse ES cells in the undifferentiated state (Lei et al. 1996; Li et al. 1992; Okano et al. 1999; Tsumura et al. 2006). However, *Dnmt1*$^{-/-}$ and *Dnmt3a/3b* DKO ES cells die upon induction of differentiation (Chen et al. 2003; Lei et al. 1996; Tucker et al. 1996). Interestingly, a recent study showed that, in contrast to mouse ES cells, human ES cells require *DNMT1*, but not *DNMT3A* and *DNMT3B*, for survival (Liao et al. 2015). It is well established that mouse and human ES cells represent different pluripotent states, with human ES cells resembling the more mature epiblast state (Tesar et al. 2007), which may explain the sensitivity of human ES cells to loss of DNA methylation. During development, the effects of DNA methylation deficiency become apparent during or after gastrulation, when the embryo differentiates to form the three germ layers (Lei et al. 1996; Li et al. 1992; Okano et al. 1999). Conditional inactivation of *Dnmt1* in mouse embryonic fibroblasts (MEFs) leads to severe hypomethylation and cell death, and *Dnmt3b*-deficient MEFs show modest hypomethylation, chromosomal instability, and abnormal cell proliferation (Dodge et al. 2005; Jackson-Grusby et al. 2001). Furthermore, although a hypomorphic mutation affecting the N-terminal region of human DNMT1 has no effect on the survival and proliferation of the colon cancer cell line HCT116 (Rhee et al. 2000), disruption of the DNMT1 catalytic domain in HCT116 leads to mitotic catastrophe and cell death (Chen et al. 2007). Taken together, these results suggest crucial roles for DNA methylation in cellular differentiation and in the viability and proper functioning of differentiated cells. Deregulation of gene expression likely plays a major role in the developmental and cellular defects associated with *Dnmt* mutations.

3 Dnmts in Genomic Imprinting

In early 1980s, elegant nuclear transplantation experiments using pronuclear stage embryos showed that mouse embryos constructed to contain only maternal or paternal diploid genome complements fail to develop beyond mid-gestation. This suggested that the parental genomes are functionally nonequivalent and marked or "imprinted" differently during male and female gametogenesis (Barton et al. 1984; McGrath and Solter 1984; Surani et al. 1984). Separate experiments using chromosome translocations in mice showed that specific chromosome segments function differently depending on the parental origin (Cattanach and Kirk 1985). In early 1990s, the first murine imprinted genes, *Igf2r*, *Igf2*, and *H19*, were discovered, which are expressed only from one parental allele (Barlow et al. 1991; Bartolomei et al. 1991; DeChiara et al. 1991). To date, approximately 150 imprinted genes, which exhibit monoallelic expression strictly according to the parental origin, have been identified in mouse (http://www.mousebook.org/mousebook-catalogs/imprinting-resource), and many of them are also imprinted in human. Imprinted genes are involved in diverse biological processes, including embryonic development, placental formation, fetal and postnatal growth, and adult behavior (Frontera et al. 2008; Reik and Walter 2001). In human, altered expression of imprinted genes, due to genetic and epigenetic changes, has been linked to infertility, molar pregnancy, and various congenital disorders such as Prader-Willi syndrome, Angelman syndrome, Beckwith-Wiedemann syndrome, and Silver-Russell syndrome (Butler 2009; Tomizawa and Sasaki 2012; Walter and Paulsen 2003). Loss of imprinting (biallelic expression or silencing of imprinted genes) is also frequently observed in cancer (Jelinic and Shaw 2007).

The majority of imprinted genes are arranged in chromosomal clusters, which usually span hundreds to thousands of kilobases. Each of the imprinting clusters is controlled by an ICR, an essential regulatory sequence that contains one or more differentially methylated regions (DMRs) between the two alleles. Thus, allele-specific DNA methylation is believed to be the primary epigenetic mark that controls the monoallelic expression of imprinted genes.

The life cycle of DNA methylation imprints consists of three major steps: establishment, maintenance, and erasure (Fig. 2).

3.1 Establishment of Methylation Imprints

DNA methylation imprints are acquired in the germ line, with the majority being established during oogenesis (maternally imprinted) and only four known loci (*H19*, *Dlk1-Gtl2*, *Rasgrf1*, and *Zdbf2*) being established during spermatogenesis (paternally imprinted). Conditional deletion of *Dnmt3a* in primordial germ cells (PGCs) disrupts both maternal and paternal imprinting. Embryos from crosses between conditional *Dnmt3a* mutant females and wild-type males die around E10.5, and conditional *Dnmt3a* mutant males are sterile due to impaired spermatogenesis (Kaneda et al. 2004). *Dnmt3L* KO mice show an identical phenotype, with the exception of

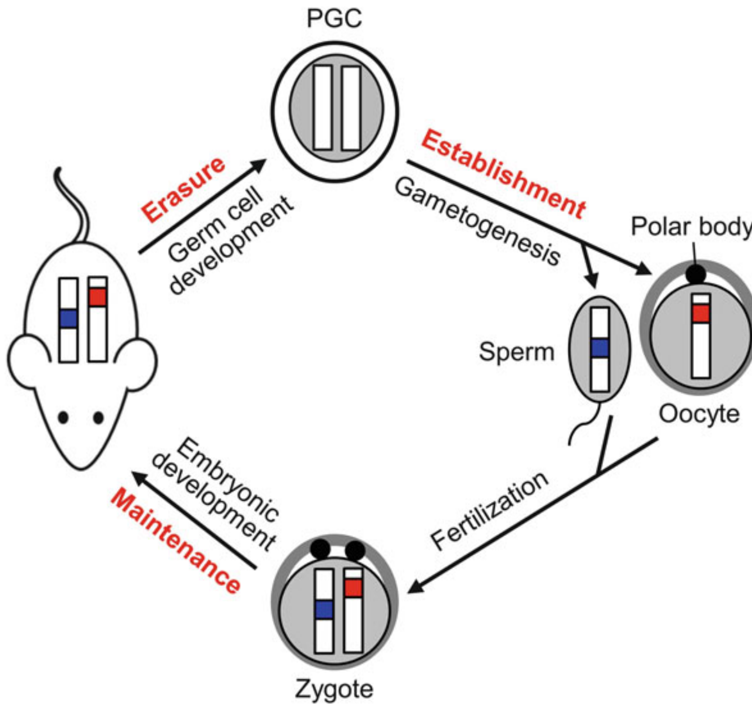


Fig. 2 Life cycle of DNA methylation imprints. The paternal (*blue*) and maternal (*red*) methylation imprints are established during gametogenesis and transmitted to the offspring through fertilization. These marks are maintained and control monoallelic expression of imprinted genes during embryogenesis and in somatic cells throughout adult life. However, they are erased in primordial germ cells (PGCs) before sex-specific methylation imprints are reestablished in later stages of germ cell development

one paternally methylated locus, *Dlk1-Gtl2*, which is methylated in *Dnmt3L* KO but not in *Dnmt3a* mutant spermatogonia (Bourc'his et al. 2001; Hata et al. 2002; Kaneda et al. 2004). In contrast, conditional deletion of *Dnmt3b* in PGCs shows no apparent phenotype (Kaneda et al. 2004). These results provide compelling genetic evidence that Dnmt3a is responsible for the establishment of germ line imprints, and Dnmt3L is an essential cofactor for Dnmt3a in this regard.

How Dnmt3L facilitates Dnmt3a function in the germ line is not fully understood. Dnmt3L, via its C-terminal domain, forms a tetrameric complex with Dnmt3a and, via its ADD domain, interacts with the N-terminal tail of histone H3 (Hata et al. 2002; Jia et al. 2007; Ooi et al. 2007; Suetake et al. 2004). These findings led to the hypothesis that Dnmt3L plays a critical role in targeting Dnmt3a to specific chromatin regions, including imprinted loci. A recent study showed that, similar to *Dnmt3L* null mutant mice, mice homozygous for an engineered point mutation (D124A) in the Dnmt3L ADD domain exhibit DNA methylation and spermatogenesis defects (Vlachogiannis et al. 2015), supporting a critical role of the ADD domain in Dnmt3L function in the male germ line. It would be interesting to

determine whether female mice homozygous for the D1124A mutation show defects in the establishment of maternal imprints. However, the Dnmt3a ADD domain also binds H3K4-unmethylated histone H3 (Otani et al. 2009; Zhang et al. 2010), which raises the question of the specific role of the Dnmt3L ADD domain. It is possible that Dnmt3L interacts with one or more proteins or histone marks that are not recognized by Dnmt3a.

The observation that H3K4 modifications disrupts the interaction between Dnmt3 proteins and histone H3 (Ooi et al. 2007; Otani et al. 2009; Zhang et al. 2010) suggests that chromatin organization may be an important determinant of the sites of *de novo* DNA methylation in the germ line. Indeed, genetic evidence indicated that the H3K4 demethylase KDM1B (also known as LSD2 and AOF1) is essential for the establishment of a subset of maternal imprints (Ciccone et al. 2009). KDM1B is highly expressed in growing oocytes, where maternal imprints are acquired, but shows little expression in most somatic tissues. *Kdm1b* KO mice are viable and show no defects in spermatogenesis and oogenesis, and male mice are fertile. However, oocytes from KDM1B-deficient females exhibit global accumulation of H3K4me2 and fail to establish DNA methylation imprints at a subset of imprinted loci. Consequently, embryos derived from these oocytes die around midgestation (Ciccone et al. 2009), similar to embryos derived from Dnmt3L- or Dnmt3a-deficient female mice (Bourc'his et al. 2001; Hata et al. 2002; Kaneda et al. 2004). These results strongly suggested that removal of H3K4me2 is a prerequisite for *de novo* DNA methylation. There is also evidence that transcription is an additional requirement in specifying DNA methylation, at least at some maternally imprinted loci. In the mouse *Gnas* locus, transcription initiated at the promoter of *Nesp55*, a gene upstream of the DMRs of the *Gnas* locus, occurs in growing oocytes, placing a large genomic region, including the DMRs, within an active transcription unit. Deletion of the *Nesp55* promoter or insertion of a transcription termination cassette downstream of *Nesp55* to ablate transcription results in failure to establish DNA methylation at the ICR of the *Gnas* locus (Chotalia et al. 2009; Frohlich et al. 2010; Williamson et al. 2011). Methylation of the DMR at the *Snrpn* locus has also been shown to depend upon transcription (Smith et al. 2011).

3.2 Maintenance of Methylation Imprints

The paternal and maternal imprints are transmitted to the zygote through fertilization, and despite extensive demethylation during preimplantation development (as described above), parental allele-specific DNA methylation imprints are faithfully maintained through development and adult life. Notably, recent genome-wide DNA methylation analyses revealed far more differentially methylated loci in oocytes and sperm than the number of imprinted genes (Kobayashi et al. 2012; Smallwood et al. 2011; Smith et al. 2012). Thus, the previous notion that imprinted loci are determined by distinct methylation patterns in gametes has been revised to the current view that genomic imprinting results from selective maintenance of germ line-derived allele-specific methylation. Genetic studies using conditional KO mice

demonstrated that Dnmt1, but not Dnmt3a or Dnmt3b, is responsible for maintaining methylation marks at imprinted loci during preimplantation development (Hirasawa et al. 2008). The oocyte-specific variant, Dnmt1o, is the predominant Dnmt1 isoform in preimplantation embryos (Hirasawa et al. 2008; Kurihara et al. 2008). However, offspring of females lacking Dnmt1o exhibit only a ~50 % reduction of methylation at certain imprinted loci (Howell et al. 2001). While initial evidence suggested that the somatic form, Dnmt1s, does not express until the blastocyst stage (Ratnam et al. 2002), subsequent work showed that Dnmt1s is present at very low levels in the nucleus of oocytes and preimplantation embryos (Hirasawa et al. 2008; Kurihara et al. 2008). Conditional deletion of *Dnmt1* (both Dnmt1o and Dnmt1s) in growing oocytes leads to a partial loss of methylation imprints in the offspring (Hirasawa et al. 2008), resembling the effect of Dnmt1o loss (Howell et al. 2001). However, ablation of both maternal and zygotic Dnmt1 results in a complete loss of methylation at paternally and maternally methylated DMRs in embryos (Hirasawa et al. 2008). Therefore, both maternal and zygotic Dnmt1 proteins are necessary for the maintenance of methylation imprints in preimplantation embryos. Dnmt1 is also responsible for the maintenance of methylation imprints in postimplantation embryos (Li et al. 1993) and likely in adult somatic cells as well.

It is not well understood what confers the specificity of Dnmt1, such that methylation is maintained at imprinted genes but not at other sequences in preimplantation embryos. Genetic and epigenetic features may distinguish imprinted loci from other regions. Several other factors have been shown to be essential for the maintenance of DNA methylation imprints. PGC7 (also known as Stella and Dppa3), a DNA-binding protein, is highly expressed in oocytes and persists in preimplantation embryos. Genetic evidence suggested that, in early embryos, maternal PGC7 plays a crucial role in protecting the maternal genome against DNA demethylation. PGC7 also protects the paternally imprinted *H19* and *Rasgrf1* against demethylation (Nakamura et al. 2007). While the mechanisms involved remain to be determined, PGC7 has been shown to play a role in chromatin condensation during oogenesis and to protect the maternal genome against Tet3-mediated conversion of 5mC to 5hmC in early embryos (Bian and Yu 2014; Liu et al. 2012; Nakamura et al. 2012). Gene targeting experiments in mice have also implicated the involvement of the *Krüppel*-associated box (KRAB)-containing zinc-finger protein ZFP57 in the maintenance of genomic imprints (Li et al. 2008), and human ZFP57 mutations are associated with hypomethylation at multiple imprinted loci in patients with transient neonatal diabetes (Mackay et al. 2008). ZFP57 specifically binds the methylated allele of ICRs, recognizing a hexanucleotide sequence (TGCCGC) shared by all murine ICRs and some human ICRs (Quenneville et al. 2011). ZFP57 interacts with KRAB-associated protein 1 (KAP1, also known as TRIM28), which acts as a scaffold protein for various heterochromatin proteins, including heterochromatin protein 1 (HP1), the histone H3K9 methyltransferase Setdb1 (also known as ESET and KMT1E), the nuclear remodeling and histone deacetylation (NuRD) complex, and Dnmt proteins and Uhrf1 (Nielsen et al. 1999; Quenneville et al. 2011; Ryan et al. 1999; Schultz et al. 2001, 2002; Zuo et al. 2012). Ablation of either maternal or zygotic KAP1 causes partial loss of DNA

methylation imprints, and ablation of both maternal and zygotic KAP1 leads to a complete loss of imprinting (Lorthongpanich et al. 2013; Messerschmidt et al. 2012; Quenneville et al. 2011). Depletion of the NuRD components methyl CpG-binding domain protein-3 (MBD3) or metastasis tumor antigen 2 (MTA2) also results in reduction of methylation at some imprinted loci in preimplantation embryos (Ma et al. 2010; Reese et al. 2007). The current view is that the ZFP57/KAP1 complex specifically recruits the DNA methylation machinery, as well as other heterochromatin proteins, to the methylated allele of ICRs to maintain genomic imprints and control monoallelic expression of imprinted genes.

3.3 Erasure of Methylation Imprints

The last step of the imprint life cycle is the erasure of methylation imprints in PGCs, which ensures the establishment of sex-specific imprints in later stages of germ cell development. In mice, PGCs are specified around E7.25 in the epiblast of the developing embryo. Shortly afterwards, PGCs begin migrating along the embryonic-extraembryonic interface and eventually arrive at the genital ridge by E12.5. Recent genome-wide DNA methylation analyses reveal that PGCs undergo demethylation in two major phases (Guibert et al. 2012; Kobayashi et al. 2013; Popp et al. 2010; Seisenberger et al. 2012). The first phase takes place during PGC expansion and migration from ~ E8.5, which leads to a global demethylation affecting almost all genomic regions. Passive demethylation likely plays a major role in this phase, as Dnmt3a, Dnmt3b, and Uhrf1 are repressed in PGCs (Kagiwada et al. 2013; Kurimoto et al. 2008). The second phase occurs from E9.5 to E13.5 and affects specific loci including ICRs, germ line-specific genes, and CpG islands on the X chromosome (Guibert et al. 2012; Hackett et al. 2013; Popp et al. 2010; Seisenberger et al. 2012; Yamaguchi et al. 2013). Genetic studies suggested that Tet1- and Tet2-mediated 5mC oxidation is important in the second phase of demethylation (Zhao and Chen 2013).

4 Dnmts in X-Chromosome Inactivation

In mammals, sex determination is controlled by a pair of sex chromosomes, the X and the Y. Whereas the Y chromosome harbors very few protein-coding genes compared to other chromosomes, the X chromosome has a relatively high gene density. The balance of X-linked gene dosage between XX females and XY males is achieved through X-chromosome inactivation (XCI), whereby one of the two X chromosomes present in female mammals is inactivated. Marsupial mammals show imprinted XCI, with only the paternal X chromosome (Xp) being inactivated. Eutherian mammals exhibit two forms of XCI: imprinted Xp inactivation in the early embryo and extraembryonic tissues and random inactivation of either Xp or the maternal X chromosome (Xm) in the embryonic (epiblast) lineage (Payer and Lee 2008; Wutz 2011).

Our knowledge about XCI mostly came from studies in the mouse. In the female zygote, both X chromosomes appear active. Soon thereafter, a series of events result in the inactivation of Xp. This imprinted XCI occurs in a two-step manner, with Xp repeat elements first silenced at the two-cell stage followed by Xp genic silencing emerging at the eight- to sixteen-cell stage. Imprinted XCI is complete by the blastocyst stage. Whereas inactivation of Xp is maintained in extraembryonic tissues, it is reversed in the ICM of the blastocyst, resulting in the biallelic expression of X-linked genes. Shortly after implantation, epiblast cells undergo random XCI (Payer and Lee 2008). Murine ES cells, derived from the ICM, represent a useful model for the study of XCI. Undifferentiated murine ES cells, like ICM cells, have two active X chromosomes, and random XCI can be recapitulated during *in vitro* differentiation (Chaumeil et al. 2004).

The process of XCI, which converts an X chromosome from relatively open euchromatin to highly condensed heterochromatin (known as the "Barr body"), can be divided into three steps: initiation of X inactivation, spreading of heterochromatin to the entire chromosome, and maintenance of the inactive state (Fig. 3). XCI is controlled by the X-inactivation center (*Xic*), a complex locus on the X chromosome that determines how many (counting step) and which X chromosomes (choice step) will be silenced. A critical gene in *Xic* encodes the "X-inactive-specific transcript" (*Xist*), a 17-kb long, non-coding RNA (lncRNA). *Xist* is expressed only from the presumptive inactive X chromosome (Xi) and then coats the same chromosome in *cis* (Clemson et al. 1996). This step is necessary and sufficient for the initiation of XCI, as targeted disruption of the *Xist* gene abrogates XCI (Marahrens et al. 1997; Penny et al. 1996), and *Xist* transgenes on autosomes can induce autosomal gene inactivation (Jiang et al. 2013; Lee and Jaenisch 1997; Lee et al. 1996). The *Xic* also harbors several other genes encoding proteins and non-coding RNAs, including the *Xist* antisense RNA *Tsix*, the *Jpx* and *Ftx* RNAs, and the E3 ubiquitin ligase RNF12, that act as part of a sophisticated regulatory network to modulate *Xist* expression in *cis* and in trans (Gendrel and Heard 2014). *Xist* is also required for the spreading of XCI from the *Xic* to the rest of the chromosome. *Xist* RNA is able to recruit PRC2, a complex responsible for the deposition of H3K27me3, which contributes to chromatin and transcriptional changes during the initiation and spreading of XCI (Zhao et al. 2008; da Rocha et al. 2014; Cifuentes-Rojas et al. 2014). Once established, the globally silent state and heterochromatin structure of Xi are transmitted through somatic cell division and clonally inherited. Although *Xist* is not required for the maintenance of gene silencing on the Xi (Brown and Willard 1994; Csankovszki et al. 1999), it appears to be important for maintaining the heterochromatin structure of Xi, as deletion of *Xist* leads to refolding of the Xi into a structure resembling the active X chromosome (Xa) (Splinter et al. 2011).

The link between DNA methylation and XCI has been well established. In somatic tissues, the 5′ end of the *Xist* gene is fully methylated on Xa and completely unmethylated on Xi. Similarly, in tissues that undergo imprinted Xp inactivation, the paternal *Xist* allele is unmethylated, and the maternal allele is fully methylated (Norris et al. 1994). Studies of *Dnmt1*-deficient ES cells and embryos revealed that XCI can occur in the absence of DNA methylation, but maintenance of *Xist*

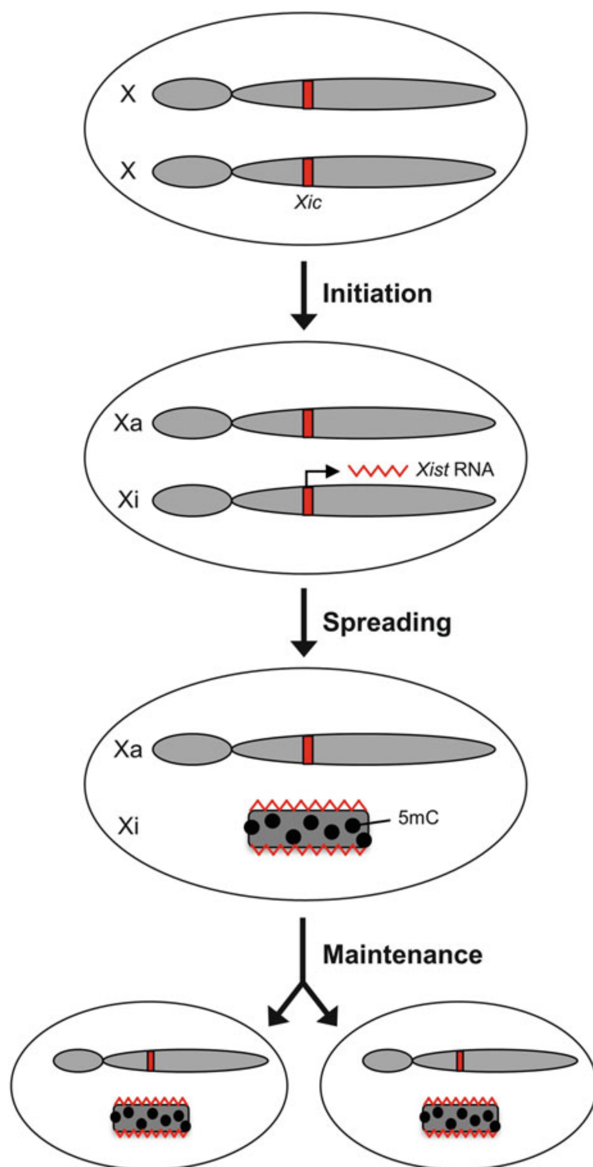


Fig. 3 Major steps of X-chromosome inactivation. The process of X-chromosome inactivation can be divided into three steps: (1) initiation (*Xist* RNA is expressed from the presumptive inactive X (Xi), but not the active X (Xa)), (2) spreading (*Xist* RNA coats the entire Xi chromosome, which recruits other factors (e.g., PRC2 complex, Dnmts) to induce heterochromatinization), and (3) maintenance (the highly compacted chromatin structure and most of the genes on Xi are stably maintained and clonally transmitted through somatic cell divisions). DNA methylation is required for the stable maintenance of Xi-linked gene silencing

promoter methylation is necessary for its stable repression in differentiated cells (Beard et al. 1995; Panning and Jaenisch 1996; Sado et al. 2000). A recent study showed that loss of Dnmt1o disrupts imprinted XCI and accentuates placental defects in females (McGraw et al. 2013). *De novo* DNA methylation is also dispensable for the initiation and propagation of XCI, as *Xist* expression is appropriately regulated and XCI occurs properly in female embryos deficient for both *Dnmt3a* and *Dnmt3b* (Sado et al. 2004). Interestingly, despite multiple mechanisms involved

in X-chromosome gene silencing, approximately 25 % of genes on Xi escape inactivation to some extent and exhibit biallelic expression in females (Carrel and Willard 2005; Yang et al. 2010). The promoter regions of these escapee genes are unmethylated (Weber et al. 2007). Furthermore, treatment of cells with the demethylating agents 5-azacytidine and 5-azadeoxycytidine has been shown to reactivate some genes on Xi (Haaf 1995). Collectively, these findings indicate that DNA methylation is not required for the initiation and propagation of XCI but is an essential component of the epigenetic mechanisms that stably maintain the silent state of Xi-linked genes.

5 Concluding Remarks

Since the discovery of mammalian Dnmts (Bestor et al. 1988; Okano et al. 1998a), great progress has been made in understanding the biological functions of DNA methylation in mammals. Genetic studies using *Dnmt* mutant mice and murine cells have provided important insights into the roles of DNA methylation in various developmental and cellular processes (Table 1). It is generally believed that DNA methylation, a relatively stable epigenetic mark, acts in concert with other epigenetic mechanisms such as histone modifications to stably maintain gene silencing and chromatin structure. It is well documented that aberrant DNA methylation patterns are associated with various human diseases. Studies in recent years have also identified genetic alterations affecting major components of the DNA methylation machinery, including DNA methylation “writers” (DNMTs), “erasers” (e.g., TETs), and “readers” (e.g., MeCP2), in cancer and developmental disorders (Hamidi et al. 2015). For example, *DNMT1* mutations are reported in two related neurodegenerative diseases (hereditary sensory and autonomic neuropathy with dementia and hearing loss type IE (HSAN IE), autosomal dominant cerebellar ataxia, deafness, and narcolepsy (ADCA-DN)), *DNMT3A* mutations are frequently found in acute myeloid leukemia (AML) and other hematologic malignancies, and *DNMT3B* mutations cause the immunodeficiency, centromeric instability, and facial anomalies (ICF) syndrome (Hamidi et al. 2015). The mechanisms by which these mutations contribute to the disease phenotypes are generally not well understood. Besides their values in elucidating the fundamental functions of DNA methylation, *Dnmt* mutant mice and cells provide important research tools for investigating the effects of *DNMT* mutations found in human patients. For instance, by expressing a Dnmt3a protein harboring a point mutation equivalent to human DNMT3A:R882H (the most prevalent *DNMT3A* mutation in AML) in *Dnmt3a* and *Dnmt3b* mutant murine ES cells, we recently demonstrated that this mutation, which occurs on only one allele in AML patients, not only leads to haploinsufficiency of DNMT3A enzymatic activity but also exhibits dominant-negative effect by forming functionally deficient complexes with wild-type DNMT3A and DNMT3B (Kim et al. 2013). Most of the *DNMT* mutations identified in patients are not null alleles, making *Dnmt* KO mice less ideal for modeling human diseases. With the development of new technologies such as CRISPR/Cas9-mediated gene editing, it now becomes more feasible to create genetically engineered animal

Table 1 Developmental phenotypes of *Dnmt* knockout mice

Gene	Mutations	Major developmental phenotypes	References
Dnmt1	*Dnmt1*$^{n/n}$	~70 % reduction in global DNA methylation, embryonic lethality at E12.5–15.5	Li et al. (1992)
	Dnmt1$^{c/c}$	~90 % reduction in global DNA methylation, developmental arrest at E8.5, embryonic lethality before 8-somite stage at ~ E9.5. Unstable random XCI	Lei et al. (1996); Sado et al. (2000)
	Dnmt1$^{s/s}$	~90 % reduction in global DNA methylation, developmental arrest at E8.5, embryonic lethality at ~ E9.5. Loss of methylation at *Xist* locus and abnormal *Xist* expression in male embryos	Beard et al. (1995); Lei et al. (1996)
	Dnmt1o$^{-/-}$	Maternal-effect phenotype: partial loss of DNA methylation imprints, defects in imprinted XCI, embryonic lethality at mid-gestation	Howell et al. (2001); McGraw et al. (2013)
	Maternal and zygotic *Dnmt3a*$^{-/-}$	Complete loss of paternal and maternal methylation imprints, embryonic lethality at mid-gestation	Hirasawa et al. (2008)
Dnmt3a	*Dnmt3a*$^{-/-}$	Gut malfunction, spermatogenesis defects, death at ~4 weeks of age	Okano et al. (1999)
	Dnmt3a$^{-/-}$ in PGCs	Failure to establish maternal and paternal methylation imprints, spermatogenesis defects	Kaneda et al. (2004)
Dnmt3b	*Dnmt3b*$^{-/-}$	Hypomethylation of minor satellite DNA, neural tube defects, embryonic lethality at E14.5–18.5	Okano et al. (1999)
	Dnmt3a$^{-/-}$, *Dnmt3b*$^{-/-}$	Failure to initiate *de novo* methylation after implantation, developmental arrest at E8.5	Okano et al. (1999)
Dnmt3L	*Dnmt3L*$^{-/-}$	Failure to establish maternal and paternal methylation imprints, spermatogenesis defects	Bourc'his et al. (2001); Hata et al. (2002)

and cellular models that better recapitulate the major features of human diseases associated with *DNMT* mutations. Genomic, epigenomic, transcriptomic, and proteomic analyses of these models will be powerful approaches for defining the molecular mechanisms and pathways involved in pathogenesis. Ultimately, such studies will likely lead to novel therapeutic and preventive strategies.

Acknowledgments Work in the Chen laboratory is supported by a Rising Star Award from Cancer Prevention and Research Institute of Texas (CPRIT, R1108) and a grant from the National Institutes of Health (NIH, 1R01DK106418-01).

References

Aapola U, Kawasaki K, Scott HS, Ollila J, Vihinen M, Heino M, Shintani A, Minoshima S, Krohn K, Antonarakis SE, et al. Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. Genomics. 2000;65:293–8.

Aapola U, Lyle R, Krohn K, Antonarakis SE, Peterson P. Isolation and initial characterization of the mouse Dnmt3l gene. Cytogenet Cell Genet. 2001;92:122–6.

Arita K, Ariyoshi M, Tochio H, Nakamura Y, Shirakawa M. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature. 2008;455:818–21.

Avvakumov GV, Walker JR, Xue S, Li Y, Duan S, Bronner C, Arrowsmith CH, Dhe-Paganon S. Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. Nature. 2008;455:822–5.

Barlow DP, Stoger R, Herrmann BG, Saito K, Schweifer N. The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. Nature. 1991;349:84–7.

Bartolomei MS, Zemel S, Tilghman SM. Parental imprinting of the mouse H19 gene. Nature. 1991;351:153–5.

Barton SC, Surani MA, Norris ML. Role of paternal and maternal genomes in mouse development. Nature. 1984;311:374–6.

Baubec T, Colombo DF, Wirbelauer C, Schmidt J, Burger L, Krebs AR, Akalin A, Schubeler D. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature. 2015;520:243–7.

Beard C, Li E, Jaenisch R. Loss of methylation activates Xist in somatic but not in embryonic cells. Genes Dev. 1995;9:2325–34.

Bestor T, Laudano A, Mattaliano R, Ingram V. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. J Mol Biol. 1988;203:971–83.

Bestor TH. Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. EMBO J. 1992;11:2611–7.

Bian C, Yu X. PGC7 suppresses TET3 for protecting DNA methylation. Nucleic Acids Res. 2014;42:2893–905.

Borgel J, Guibert S, Li Y, Chiba H, Schubeler D, Sasaki H, Forne T, Weber M. Targets and dynamics of promoter DNA methylation during early mouse development. Nat Genet. 2010;42:1093–100.

Bostick M, Kim JK, Esteve PO, Clark A, Pradhan S, Jacobsen SE. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science. 2007;317:1760–4.

Bourc'his D, Bestor TH. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature. 2004;431:96–9.

Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH. Dnmt3L and the establishment of maternal genomic imprints. Science. 2001;294:2536–9.

Brown CJ, Willard HF. The human X-inactivation centre is not required for maintenance of X-chromosome inactivation. Nature. 1994;368:154–6.

Butler MG. Genomic imprinting disorders in humans: a mini-review. J Assist Reprod Genet. 2009;26:477–86.

Carrel L, Willard HF. X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature. 2005;434:400–4.

Cattanach BM, Kirk M. Differential activity of maternally and paternally derived chromosome regions in mice. Nature. 1985;315:496–8.

Chaumeil J, Okamoto I, Heard E. X-chromosome inactivation in mouse embryonic stem cells: analysis of histone modifications and transcriptional activity using immunofluorescence and FISH. Methods Enzymol. 2004;376:405–19.

Chen T, Hevi S, Gay F, Tsujimoto N, He T, Zhang B, Ueda Y, Li E. Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. Nat Genet. 2007;39:391–6.

Chen T, Tsujimoto N, Li E. The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. Mol Cell Biol. 2004;24:9048–58.

Chen T, Ueda Y, Dodge JE, Wang Z, Li E. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. Mol Cell Biol. 2003;23:5594–605.

Chen T, Ueda Y, Xie S, Li E. A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active de novo methylation. J Biol Chem. 2002;277:38746–54.

Cheng J, Yang Y, Fang J, Xiao J, Zhu T, Chen F, Wang P, Li Z, Yang H, Xu Y. Structural insight into coordinated recognition of trimethylated histone H3 lysine 9 (H3K9me3) by the plant homeodomain (PHD) and tandem tudor domain (TTD) of UHRF1 (ubiquitin-like, containing PHD and RING finger domains, 1) protein. J Biol Chem. 2013;288:1329–39.

Chotalia M, Smallwood SA, Ruf N, Dawson C, Lucifero D, Frontera M, James K, Dean W, Kelsey G. Transcription is required for establishment of germline methylation marks at imprinted genes. Genes Dev. 2009;23:105–17.

Chuang LS, Ian HI, Koh TW, Ng HH, Xu G, Li BF. Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. Science. 1997;277:1996–2000.

Ciccone DN, Su H, Hevi S, Gay F, Lei H, Bajko J, Xu G, Li E, Chen T. KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. Nature. 2009;461:415–8.

Cifuentes-Rojas C, Hernandez AJ, Sarma K, Lee JT. Regulatory interactions between RNA and polycomb repressive complex 2. Mol Cell. 2014;55:171–85.

Clemson CM, McNeil JA, Willard HF, Lawrence JB. XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. J Biol Chem. 1996;132:259–75.

Csankovszki G, Panning B, Bates B, Pehrson JR, Jaenisch R. Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X inactivation. Nat Genet. 1999;22:323–4.

da Rocha ST, Boeva V, Escamilla-Del-Arenal M, Ancelin K, Granier C, Matias NR, Sanulli S, Chow J, Schulz E, Picard C, et al. Jarid2 is implicated in the initial xist-induced targeting of PRC2 to the inactive X chromosome. Mol Cell. 2014;53:301–16.

DeChiara TM, Robertson EJ, Efstratiadis A. Parental imprinting of the mouse insulin-like growth factor II gene. Cell. 1991;64:849–59.

Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S, Jeltsch A. The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. J Biol Chem. 2010;285:26114–20.

Ding F, Chaillet JR. In vivo stabilization of the Dnmt1 (cytosine-5)- methyltransferase protein. Proc Natl Acad Sci U S A. 2002;99:14861–6.

Dodge JE, Okano M, Dick F, Tsujimoto N, Chen T, Wang S, Ueda Y, Dyson N, Li E. Inactivation of Dnmt3b in mouse embryonic fibroblasts results in DNA hypomethylation, chromosomal instability, and spontaneous immortalization. J Biol Chem. 2005;280:17986–91.

Du Z, Song J, Wang Y, Zhao Y, Guda K, Yang S, Kao HY, Xu Y, Willis J, Markowitz SD, et al. DNMT1 stability is regulated by proteins coordinating deubiquitination and acetylation-driven ubiquitination. Sci Signal. 2010;3:ra80.

Easwaran HP, Schermelleh L, Leonhardt H, Cardoso MC. Replication-independent chromatin loading of Dnmt1 during G2 and M phases. EMBO Rep. 2004;5:1181–6.

Feng J, Zhou Y, Campbell SL, Le T, Li E, Sweatt JD, Silva AJ, Fan G. Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. Nat Neurosci. 2010;13:423–30.

Frohlich LF, Mrakovcic M, Steinborn R, Chung UI, Bastepe M, Juppner H. Targeted deletion of the Nesp55 DMR defines another Gnas imprinting control region and provides a mouse model of autosomal dominant PHP-Ib. Proc Natl Acad Sci U S A. 2010;107:9275–80.

Frontera M, Dickins B, Plagge A, Kelsey G. Imprinted genes, postnatal adaptations and enduring effects on energy homeostasis. Adv Exp Med Biol. 2008;626:41–61.

Gendrel AV, Heard E. Noncoding RNAs and epigenetic mechanisms during X-chromosome inactivation. Annu Rev Cell Dev Biol. 2014;30:561–80.
Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh CL, Zhang X, Golic KG, Jacobsen SE, Bestor TH. Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science. 2006;311:395–8.
Gu TP, Guo F, Yang H, Wu HP, Xu GF, Liu W, Xie ZG, Shi L, He X, Jin SG, et al. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature. 2011;477:606–10.
Guibert S, Forne T, Weber M. Global profiling of DNA methylation erasure in mouse primordial germ cells. Genome Res. 2012;22:633–41.
Haaf T. The effects of 5-azacytidine and 5-azadeoxycytidine on chromosome structure and function: implications for methylation-associated cellular processes. Pharmacol Ther. 1995;65:19–46.
Hackett JA, Sengupta R, Zylicz JJ, Murakami K, Lee C, Down TA, Surani MA. Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. Science. 2013;339:448–52.
Hamidi T, Singh AK, Chen T. Genetic alterations of DNA methylation machinery in human diseases. Epigenomics. 2015;7:247–65.
Hashimoto H, Horton JR, Zhang X, Bostick M, Jacobsen SE, Cheng X. The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. Nature. 2008;455:826–9.
Hata K, Okano M, Lei H, Li E. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Development. 2002;129:1983–93.
He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science. 2011;333:1303–7.
Hirasawa R, Chiba H, Kaneda M, Tajima S, Li E, Jaenisch R, Sasaki H. Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. Genes Dev. 2008;22:1607–16.
Hirasawa R, Sasaki H. Dynamic transition of Dnmt3b expression in mouse pre- and early post-implantation embryos. Gene Expr Patterns. 2009;9:27–30.
Holliday R, Pugh JE. DNA modification mechanisms and gene activity during development. Science. 1975;187:226–32.
Howell CY, Bestor TH, Ding F, Latham KE, Mertineit C, Trasler JM, Chaillet JR. Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. Cell. 2001;104:829–38.
Inoue A, Shen L, Dai Q, He C, Zhang Y. Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. Cell Res. 2011;21:1670–6.
Inoue A, Zhang Y. Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. Science. 2011;334:194.
Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science. 2011;333:1300–3.
Jackson-Grusby L, Beard C, Possemato R, Tudor M, Fambrough D, Csankovszki G, Dausman J, Lee P, Wilson C, Lander E, et al. Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nat Genet. 2001;27:31–9.
Jelinic P, Shaw P. Loss of imprinting and cancer. J Pathol. 2007;211:261–8.
Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature. 2007;449:248–51.
Jiang J, Jing Y, Cost GJ, Chiang JC, Kolpa HJ, Cotton AM, Carone DM, Carone BR, Shivak DA, Guschin DY, et al. Translating dosage compensation to trisomy 21. Nature. 2013;500:296–300.
Kagiwada S, Kurimoto K, Hirota T, Yamaji M, Saitou M. Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. EMBO J. 2013;32:340–53.
Kaneda M, Okano M, Hata K, Sado T, Tsujimoto N, Li E, Sasaki H. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature. 2004;429:900–3.
Kim SJ, Zhao H, Hardikar S, Singh AK, Goodell MA, Chen T. A DNMT3A mutation common in AML exhibits dominant-negative effects in murine ES cells. Blood. 2013;122:4086–9.
Kobayashi H, Sakurai T, Imai M, Takahashi N, Fukuda A, Yayoi O, Sato S, Nakabayashi K, Hata K, Sotomaru Y, et al. Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. PLoS Genet. 2012;8:e1002440.

Kobayashi H, Sakurai T, Miura F, Imai M, Mochiduki K, Yanagisawa E, Sakashita A, Wakai T, Suzuki Y, Ito T, et al. High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. Genome Res. 2013;23:616–27.
Kurihara Y, Kawamura Y, Uchijima Y, Amamo T, Kobayashi H, Asano T, Kurihara H. Maintenance of genomic methylation patterns during preimplantation development requires the somatic form of DNA methyltransferase 1. Dev Biol. 2008;313:335–46.
Kurimoto K, Yabuta Y, Ohinata Y, Shigeta M, Yamanaka K, Saitou M. Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. Genes Dev. 2008;22:1617–35.
Lee JT, Jaenisch R. Long-range cis effects of ectopic X-inactivation centres on a mouse autosome. Nature. 1997;386:275–9.
Lee JT, Strauss WM, Dausman JA, Jaenisch R. A 450 kb transgene displays properties of the mammalian X-inactivation center. Cell. 1996;86:83–94.
Lee PP, Fitzpatrick DR, Beard C, Jessup HK, Lehar S, Makar KW, Perez-Melgosa M, Sweetser MT, Schlissel MS, Nguyen S, et al. A critical role for Dnmt1 and DNA methylation in T cell development, function, and survival. Immunity. 2001;15:763–74.
Lei H, Oh SP, Okano M, Juttermann R, Goss KA, Jaenisch R, Li E. De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development. 1996;122:3195–205.
Leonhardt H, Page AW, Weier HU, Bestor TH. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell. 1992;71:865–73.
Li E, Beard C, Jaenisch R. Role for DNA methylation in genomic imprinting. Nature. 1993; 366:362–5.
Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell. 1992;69:915–26.
Li X, Ito M, Zhou F, Youngson N, Zuo X, Leder P, Ferguson-Smith AC. A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. Dev Cell. 2008;15:547–57.
Liao J, Karnik R, Gu H, Ziller MJ, Clement K, Tsankov AM, Akopian V, Gifford CA, Donaghey J, Galonska C, et al. Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. Nat Genet. 2015;47:469–78.
Liu X, Gao Q, Li P, Zhao Q, Zhang J, Li J, Koseki H, Wong J. UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. Nat Commun. 2013;4:1563.
Liu YJ, Nakamura T, Nakano T. Essential role of DPPA3 for chromatin condensation in mouse oocytogenesis. Biol Reprod. 2012;86:40.
Lorthongpanich C, Cheow LF, Balu S, Quake SR, Knowles BB, Burkholder WF, Solter D, Messerschmidt DM. Single-cell DNA-methylation analysis reveals epigenetic chimerism in preimplantation embryos. Science. 2013;341:1110–2.
Ma P, Lin S, Bartolomei MS, Schultz RM. Metastasis tumor antigen 2 (MTA2) is involved in proper imprinted expression of H19 and Peg3 during mouse preimplantation development. Biol Reprod. 2010;83:1027–35.
Mackay DJ, Callaway JL, Marks SM, White HE, Acerini CL, Boonen SE, Dayanikli P, Firth HV, Goodship JA, Haemers AP, et al. Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. Nat Genet. 2008;40: 949–51.
Marahrens Y, Panning B, Dausman J, Strauss W, Jaenisch R. Xist-deficient mice are defective in dosage compensation but not spermatogenesis. Genes Dev. 1997;11:156–66.
McGrath J, Solter D. Completion of mouse embryogenesis requires both the maternal and paternal genomes. Cell. 1984;37:179–83.
McGraw S, Oakes CC, Martel J, Cirio MC, de Zeeuw P, Mak W, Plass C, Bartolomei MS, Chaillet JR, Trasler JM. Loss of DNMT1o disrupts imprinted X chromosome inactivation and accentuates placental defects in females. PLoS Genet. 2013;9:e1003873.
Mertineit C, Yoder JA, Taketo T, Laird DW, Trasler JM, Bestor TH. Sex-specific exons control DNA methyltransferase in mammalian germ cells. Development. 1998;125:889–97.

Messerschmidt DM, de Vries W, Ito M, Solter D, Ferguson-Smith A, Knowles BB. Trim28 is required for epigenetic stability during mouse oocyte to embryo transition. Science. 2012;335:1499–502.

Mudbhary R, Hoshida Y, Chernyavskaya Y, Jacob V, Villanueva A, Fiel MI, Chen X, Kojima K, Thung S, Bronson RT, et al. UHRF1 overexpression drives DNA hypomethylation and hepatocellular carcinoma. Cancer Cell. 2014;25:196–209.

Muto M, Kanari Y, Kubo E, Takabe T, Kurihara T, Fujimori A, Tatsumi K. Targeted disruption of Np95 gene renders murine embryonic stem cells hypersensitive to DNA damaging agents and DNA replication blocks. J Biol Chem. 2002;277:34549–55.

Nakamura T, Arai Y, Umehara H, Masuhara M, Kimura T, Taniguchi H, Sekimoto T, Ikawa M, Yoneda Y, Okabe M, et al. PGC7/Stella protects against DNA demethylation in early embryogenesis. Nat Cell Biol. 2007;9:64–71.

Nakamura T, Liu YJ, Nakashima H, Umehara H, Inoue K, Matoba S, Tachibana M, Ogura A, Shinkai Y, Nakano T. PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. Nature. 2012;486:415–9.

Neri F, Krepelova A, Incarnato D, Maldotti M, Parlato C, Galvagni F, Matarese F, Stunnenberg HG, Oliviero S. Dnmt3L antagonizes DNA methylation at bivalent promoters and favors DNA methylation at gene bodies in ESCs. Cell. 2013;155:121–34.

Nielsen AL, Ortiz JA, You J, Oulad-Abdelghani M, Khechumian R, Gansmuller A, Chambon P, Losson R. Interaction with members of the heterochromatin protein 1 (HP1) family and histone deacetylation are differentially involved in transcriptional silencing by members of the TIF1 family. EMBO J. 1999;18:6385–95.

Nishiyama A, Yamaguchi L, Sharif J, Johmura Y, Kawamura T, Nakanishi K, Shimamura S, Arita K, Kodama T, Ishikawa F, et al. Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. Nature. 2013;502:249–53.

Noh KM, Wang H, Kim HR, Wenderski W, Fang F, Li CH, Dewell S, Hughes SH, Melnick AM, Patel DJ, et al. Engineering of a histone-recognition domain in Dnmt3a alters the epigenetic landscape and phenotypic features of mouse ESCs. Mol Cell. 2015;59:89–103.

Norris DP, Patel D, Kay GF, Penny GD, Brockdorff N, Sheardown SA, Rastan S. Evidence that random and imprinted Xist expression is controlled by preemptive methylation. Cell. 1994;77:41–51.

Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell. 1999;99:247–57.

Okano M, Xie S, Li E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet. 1998a;19:219–20.

Okano M, Xie S, Li E. Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. Nucleic Acids Res. 1998b;26:2536–40.

Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin SP, Allis CD, et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature. 2007;448:714–7.

Ostler KR, Yang Q, Looney TJ, Zhang L, Vasanthakumar A, Tian Y, Kocherginsky M, Raimondi SL, DeMaio JG, Salwen HR, et al. Truncated DNMT3B isoform DNMT3B7 suppresses growth, induces differentiation, and alters DNA methylation in human neuroblastoma. Cancer Res. 2012;72:4714–23.

Otani J, Nankumo T, Arita K, Inamoto S, Ariyoshi M, Shirakawa M. Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. EMBO Rep. 2009;10:1235–41.

Panning B, Jaenisch R. DNA hypomethylation can activate Xist expression and silence X-linked genes. Genes Dev. 1996;10:1991–2002.

Payer B, Lee JT. X chromosome dosage compensation: how mammals keep the balance. Annu Rev Genet. 2008;42:733–72.

Penny GD, Kay GF, Sheardown SA, Rastan S, Brockdorff N. Requirement for Xist in X chromosome inactivation. Nature. 1996;379:131–7.

Popp C, Dean W, Feng S, Cokus SJ, Andrews S, Pellegrini M, Jacobsen SE, Reik W. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature. 2010;463:1101–5.

Pradhan S, Bacolla A, Wells RD, Roberts RJ. Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. J Biol Chem. 1999;274:33002–10.

Qin W, Leonhardt H, Spada F. Usp7 and Uhrf1 control ubiquitination and stability of the maintenance DNA methyltransferase Dnmt1. J Cell Biochem. 2011;112:439–44.

Qin W, Wolf P, Liu N, Link S, Smets M, Mastra FL, Forne I, Pichler G, Horl D, Fellinger K, et al. DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination. Cell Res. 2015;25:911–29.

Quenneville S, Verde G, Corsinotti A, Kapopoulou A, Jakobsson J, Offner S, Baglivo I, Pedone PV, Grimaldi G, Riccio A, et al. In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. Mol Cell. 2011;44:361–72.

Ratnam S, Mertineit C, Ding F, Howell CY, Clarke HJ, Bestor TH, Chaillet JR, Trasler JM. Dynamics of Dnmt1 methyltransferase expression and intracellular localization during oogenesis and preimplantation development. Dev Biol. 2002;245:304–14.

Reese KJ, Lin S, Verona RI, Schultz RM, Bartolomei MS. Maintenance of paternal methylation and repression of the imprinted H19 gene requires MBD3. PLoS Genet. 2007;3:e137.

Reik W, Walter J. Genomic imprinting: parental influence on the genome. Nat Rev Genet. 2001;2:21–32.

Rhee I, Jair KW, Yen RW, Lengauer C, Herman JG, Kinzler KW, Vogelstein B, Baylin SB, Schuebel KE. CpG methylation is maintained in human cancer cells lacking DNMT1. Nature. 2000;404:1003–7.

Riggs AD. X inactivation, differentiation, and DNA methylation. Cytogenet Cell Genet. 1975; 14:9–25.

Rothbart SB, Dickson BM, Ong MS, Krajewski K, Houliston S, Kireev DB, Arrowsmith CH, Strahl BD. Multivalent histone engagement by the linked tandem Tudor and PHD domains of UHRF1 is required for the epigenetic inheritance of DNA methylation. Genes Dev. 2013;27:1288–98.

Rothbart SB, Krajewski K, Nady N, Tempel W, Xue S, Badeaux AI, Barsyte-Lovejoy D, Martinez JY, Bedford MT, Fuchs SM, et al. Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. Nat Struct Mol Biol. 2012;19:1155–60.

Rottach A, Frauer C, Pichler G, Bonapace IM, Spada F, Leonhardt H. The multi-domain protein Np95 connects DNA methylation and histone modification. Nucleic Acids Res. 2010;38:1796–804.

Rouleau J, Tanigawa G, Szyf M. The mouse DNA methyltransferase 5′-region. A unique housekeeping gene promoter. J Biol Chem. 1992;267:7368–77.

Ryan RF, Schultz DC, Ayyanathan K, Singh PB, Friedman JR, Fredericks WJ, Rauscher 3rd FJ. KAP-1 corepressor protein interacts and colocalizes with heterochromatic and euchromatic HP1 proteins: a potential role for Kruppel-associated box-zinc finger proteins in heterochromatin-mediated gene silencing. Mol Cell Biol. 1999;19:4366–78.

Sado T, Fenner MH, Tan SS, Tam P, Shioda T, Li E. X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. Dev Biol. 2000;225:294–303.

Sado T, Okano M, Li E, Sasaki H. De novo DNA methylation is dispensable for the initiation and propagation of X chromosome inactivation. Development. 2004;131:975–82.

Schneider K, Fuchs C, Dobay A, Rottach A, Qin W, Wolf P, Alvarez-Castro JM, Nalaskowski MM, Kremmer E, Schmid V, et al. Dissection of cell cycle-dependent dynamics of Dnmt1 by FRAP and diffusion-coupled modeling. Nucleic Acids Res. 2013;41:4860–76.

Schultz DC, Ayyanathan K, Negorev D, Maul GG, Rauscher 3rd FJ. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. Genes Dev. 2002;16:919–32.

Schultz DC, Friedman JR, Rauscher 3rd FJ. Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2alpha subunit of NuRD. Genes Dev. 2001;15:428–43.
Seisenberger S, Andrews S, Krueger F, Arand J, Walter J, Santos F, Popp C, Thienpont B, Dean W, Reik W. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol Cell. 2012;48:849–62.
Shah MY, Vasanthakumar A, Barnes NY, Figueroa ME, Kamp A, Hendrick C, Ostler KR, Davis EM, Lin S, Anastasi J, et al. DNMT3B7, a truncated DNMT3B isoform expressed in human tumors, disrupts embryonic development and accelerates lymphomagenesis. Cancer Res. 2010;70:5840–50.
Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, Endo TA, Shinga J, Mizutani-Koseki Y, Toyoda T, Okamura K, et al. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature. 2007;450:908–12.
Smallwood SA, Tomizawa S, Krueger F, Ruf N, Carli N, Segonds-Pichon A, Sato S, Hata K, Andrews SR, Kelsey G. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat Genet. 2011;43:811–4.
Smith EY, Futtner CR, Chamberlain SJ, Johnstone KA, Resnick JL. Transcription is required to establish maternal imprinting at the Prader-Willi syndrome and Angelman syndrome locus. PLoS Genet. 2011;7:e1002422.
Smith ZD, Chan MM, Mikkelsen TS, Gu H, Gnirke A, Regev A, Meissner A. A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature. 2012;484:339–44.
Splinter E, de Wit E, Nora EP, Klous P, van de Werken HJ, Zhu Y, Kaaij LJ, van Ijcken W, Gribnau J, Heard E, et al. The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. Genes Dev. 2011;25:1371–83.
Suetake I, Shinozaki F, Miyagawa J, Takeshima H, Tajima S. DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. J Biol Chem. 2004;279:27816–23.
Surani MA, Barton SC, Norris ML. Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. Nature. 1984;308:548–50.
Takebayashi S, Tamura T, Matsuoka C, Okano M. Major and essential role for the DNA methylation mark in mouse embryogenesis and stable association of DNMT1 with newly replicated regions. Mol Cell Biol. 2007;27:8243–58.
Takeshita K, Suetake I, Yamashita E, Suga M, Narita H, Nakagawa A, Tajima S. Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). Proc Natl Acad Sci U S A. 2011;108:9055–9.
Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, Mack DL, Gardner RL, McKay RD. New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature. 2007;448:196–9.
Tomizawa S, Sasaki H. Genomic imprinting and its relevance to congenital disease, infertility, molar pregnancy and induced pluripotent stem cell. J Hum Genet. 2012;57:84–91.
Tsumura A, Hayakawa T, Kumaki Y, Takebayashi S, Sakaue M, Matsuoka C, Shimotohno K, Ishikawa F, Li E, Ueda HR, et al. Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. Genes Cells. 2006;11:805–14.
Tucker KL, Talbot D, Lee MA, Leonhardt H, Jaenisch R. Complementation of methylation deficiency in embryonic stem cells by a DNA methyltransferase minigene. Proc Natl Acad Sci U S A. 1996;93:12920–5.
Van den Wyngaert I, Sprengel J, Kass SU, Luyten WH. Cloning and analysis of a novel human putative DNA methyltransferase. FEBS Lett. 1998;426:283–9.
Vlachogiannis G, Niederhuth CE, Tuna S, Stathopoulou A, Viiri K, de Rooij DG, Jenner RG, Schmitz RJ, Ooi SK. The Dnmt3L ADD domain controls cytosine methylation establishment during spermatogenesis. Cell Rep. 2015;10:944–56.
Walter J, Paulsen M. Imprinting and disease. Semin Cell Dev Biol. 2003;14:101–10.

Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, Schubeler D. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet. 2007;39:457–66.

Williamson CM, Ball ST, Dawson C, Mehta S, Beechey CV, Fray M, Teboul L, Dear TN, Kelsey G, Peters J. Uncoupling antisense-mediated silencing and DNA methylation in the imprinted Gnas cluster. PLoS Genet. 2011;7:e1001347.

Wossidlo M, Nakamura T, Lepikhov K, Marques CJ, Zakhartchenko V, Boiani M, Arand J, Nakano T, Reik W, Walter J. 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nat Commun. 2011;2:241.

Wutz A. Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. Nat Rev Genet. 2011;12:542–53.

Yamaguchi S, Hong K, Liu R, Inoue A, Shen L, Zhang K, Zhang Y. Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming. Cell Res. 2013;23:329–39.

Yang F, Babak T, Shendure J, Disteche CM. Global survey of escape from X inactivation by RNA-sequencing in mouse. Genome Res. 2010;20:614–22.

Yoder JA, Bestor TH. A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. Hum Mol Genet. 1998;7:279–84.

Zhang Y, Jurkowska R, Soeroes S, Rajavelu A, Dhayalan A, Bock I, Rathert P, Brandt O, Reinhardt R, Fischle W, et al. Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. Nucleic Acids Res. 2010;38:4246–53.

Zhao H, Chen T. Tet family of 5-methylcytosine dioxygenases in mammalian development. J Hum Genet. 2013;58:421–7.

Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science. 2008;322:750–6.

Zuo X, Sheng J, Lau HT, McDonald CM, Andrade M, Cullen DE, Bell FT, Iacovino M, Kyba M, Xu G, et al. Zinc finger protein ZFP57 requires its co-factor to recruit DNA methyltransferases and maintains DNA methylation imprint in embryonic stem cells via its transcriptional repression domain. J Biol Chem. 2012;287:2107–18.

The Role of DNA Methylation in Cancer

Ranjani Lakshminarasimhan and Gangning Liang

Abstract

The malignant transformation of normal cells is driven by both genetic and epigenetic changes. With the advent of next-generation sequencing and large-scale multinational consortium studies, it has become possible to profile the genomes and epigenomes of thousands of primary tumors from nearly every cancer type. From these genome-wide studies, it became clear that the dynamic regulation of DNA methylation is a critical epigenetic mechanism of cancer initiation, maintenance, and progression. Proper control of DNA methylation is not only crucial for regulating gene transcription, but its broader consequences include maintaining the integrity of the genome and modulating immune response. Here, we describe the aberrant DNA methylation changes that take place in cancer and how they contribute to the disease phenotype. Further, we highlight potential clinical implications of these changes in the context of prognostic and diagnostic biomarkers, as well as therapeutic targets.

Abbreviations

AML	Acute myeloid leukemia
CGIs	CpG islands
CIMP	CpG island methylator phenotype
CpG	Cytosine-guanine dinucleotide
DNMT	DNA methyltransferases

R. Lakshminarasimhan • G. Liang (✉)
Department of Urology, University of Southern California, Norris Comprehensive Cancer Center, Los Angeles, CA 90089, USA
e-mail: gliang@usc.edu

© Springer International Publishing Switzerland 2016

A. Jeltsch, R.Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*,
Advances in Experimental Medicine and Biology 945,
DOI 10.1007/978-3-319-43624-1_7

DNMTi	DNA methyltransferase inhibitor
dsRNA	Double-stranded RNA
ERV	Endogenous retrovirus
GBM	Glioblastoma multiforme
MDS	Myelodysplastic syndrome
TCGA	The Cancer Genome Atlas
TSGs	Tumor suppressor genes

1 Cancer and Epigenetics

Classic hallmarks of cancer, as described by Hanahan and Weinberg, include maintenance of cell proliferation, evasion of growth suppression and cell death, promotion of angiogenesis, invasion, and metastasis (Hanahan and Weinberg 2011). Both genetic and epigenetic alterations underlie these processes. Genetic changes contributing to tumorigenesis have been well studied and include missense mutations, copy number variations, insertions, deletions, and recombination of DNA. Complementary to these genetic events, it is now accepted that oncogenic traits also accumulate through epigenetic disturbances (Baylin and Jones 2011; Sandoval and Esteller 2012).

DNA methylation, histone tail modifications, nucleosome positioning, and noncoding RNA are the epigenetic mechanisms crucial for the maintenance of heritable changes in gene expression potential and chromatin organization over cell generations. Epigenetic regulation of transcription allows genetically identical cells to establish distinct cellular phenotypes.

Recent next-generation sequencing studies of cancer genomes have revealed frequent and recurrent mutations in a wide variety of epigenetic modulators, including mediators of DNA methylation, covalent histone modifiers, and genes encoding subunits of chromatin remodelers (You and Jones 2012; Shen and Laird 2013). Aberrant activity of these key epigenetic players results in the deregulation of gene expression and has been implicated in many malignancies, including numerous cancers (Sharma et al. 2010; Hanahan and Weinberg 2011).

2 DNA Methylation and DNA Methyltransferases

Mammalian DNA methylation primarily occurs as a covalent addition of methyl group to the carbon-5 atom of cytosine in a cytosine-guanine (CpG) dinucleotide. This enzymatic reaction is catalyzed by three DNA methyltransferases (DNMTs). DNMT3A and DNMT3B show equal preference to hemimethylated and unmethylated DNA molecules and are essential for the creation of initial DNA methylation marks (Okano et al. 1999). DNMT3A and DNMT3B are highly expressed in embryonic stem (ES) cells and, though downregulated, continue to be expressed in somatic cells (Sharma et al. 2011). After replication of the DNA, the newly synthesized strand does not carry the methylation modification. DNMT1 preferentially

catalyzes the covalent addition of the methyl group onto the unmethylated strand of the hemimethylated DNA molecule. While DNMT1 carries out the majority of the DNA methylation in a dividing cell, DNMT3a/3b strongly associate with nucleosomes to permit efficient propagation of DNA methylation by modification of those sites missed by DNMT1 (Okano et al. 1999; Liang et al. 2002; Rhee et al. 2002; Jones and Liang 2009; Sharma et al. 2011).

DNMTs are responsible for laying down methyl groups, whereas the recently identified ten-eleven translocation (TET) family of dioxygenases provide a paradigm for DNA demethylation. These enzymes, through successive enzymatic reactions, can oxidize 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) and 5-formylcytosine (5fC) to 5-carboxylcytosine (5caC) (Ko et al. 2010; Pastor et al. 2011, 2013). The oxidization of 5mC contributes to the passive loss of DNA methylation over cell replication. In addition, the oxidized intermediates can be restored to cytosine by iterative oxidation followed by base excision repair mediated by thymine DNA glycosylase (TDG) (Kohli and Zhang 2013). Together, with DNMTs, these enzymes provide a model for the dynamic regulation of DNA methylation.

3 CpG Islands

Methylated cytosine residues are susceptible to spontaneous deamination resulting in the poorly repaired cytosine to thymine transition. As a result, nearly a third of all disease-causing familial mutations and single-nucleotide polymorphisms are found in methylated CpG sites. Similarly, in somatic cells, CpG residues in the gene body or coding regions habitually contribute to mutational hot spots, such as in the case of inactivating C to T transitions at the tumor suppressor gene *p53* (Pfeifer 2000; Jones and Baylin 2002).

Another consequence of this phenomenon is that there is a reduced representation of the CpG palindrome globally in the human genome, except in genomic regions designated as CpG islands (CGIs). CGIs were first defined by Gardiner-Garden and Frommer as a 200-bp DNA with a C+G content of 50 % and an (observed CpG)/(expected CpG) in excess of 0.6 (Gardiner-Garden and Frommer 1987). While the majority of CpGs are methylated, CpG sites located in CGIs remain overwhelmingly unmethylated (Meissner et al. 2008). These islands are often, but not exclusively, found in the nearly half of all gene promoters (Mikkelsen et al. 2007; Meissner et al. 2008). Non-CGI promoters, on the other hand, are predominantly methylated and silent. These genes are more likely to be tissue specifically expressed; therefore, only a small subset of non-CGI promoters remain unmethylated and accessible for transcription factors in each tissue type (Eckhardt et al. 2006).

4 DNA Methylation in Normal Mammalian Tissue

Under normal physiological conditions, DNA methylation is vital to the maintenance of genome integrity, as methylation of repeat regions prevents retrotransposon activity. It is also involved in suppressing genes in a tissue-specific context and

in facilitating allelic expression through genomic imprinting, and it is required for the inactivation of the additional copy of the X-chromosome in females (Smith and Meissner 2013).

Over evolutionary times, the mammalian genome has accumulated a large number of parasitic transposable, retroviral, repeat elements. These elements make up more than a third of the human genome (Cordaux and Batzer 2009). CpG methylation of transposable elements silences the elements and prevents their transcription. DNA methylation of these repeat elements is central to the maintenance of genomic integrity (Yoder et al. 1997).

The epigenetic phenomenon of genomic imprinting results in the unequal contribution of the chromosomes inherited from each parent to the embryonic development. Imprinted genes are expressed in a parental-origin-specific manner rather than from both chromosomes. DNA methylation is the key mechanism, by which the allele-specific expression is established and maintained. For example, if the maternal allele is imprinted by DNA methylation, then it becomes silenced, and only the gene inherited from the father is expressed (Li et al. 1993; Ferguson-Smith 2011).

X-chromosome inactivation is a developmentally necessary process, by which the dosage of X-linked genes in females is equalized to the dosage of those genes in males (Pessia et al. 2012). In mammals, the choice of the X-chromosome to be inactivated is random. The process is initiated and propagated by the increased expression of the noncoding RNA *XIST* on the X-chromosome that is going to be inactivated (Xi). This then triggers a cascade of events that finally result in the exclusion of RNA polymerase, as well as the recruitment of repressive histone marks to Xi (Pontier and Gribnau 2011). Once the inactive X has been established, DNA methylation of CpG islands is necessary for the maintenance of the silenced state (Bestor et al. 2015).

5 DNA Methylation in Cancer

Broad changes of the epigenome accompany cancer initiation and progression. It has been known for decades that cancer cells display a global loss of CpG methylation, including regions with low density of CpG sites, repeat elements, retrotransposons, and laminin-associated domains (LADs). This phenomenon occurs juxtaposed with concomitant locus-specific hypermethylation at CpG islands and CpG island shores (Weisenberger and Liang 2015).

5.1 Hypermethylation

5.1.1 Promoters

Epigenetic processes such as DNA methylation serve as a secondary mechanism for the inactivation of tumor suppressor genes (TSGs) in addition to genetic changes (Jones and Laird 1999; You and Jones 2012). The hypermethylation of CGI promoters in cancer cells is inversely correlated with gene expression and results in

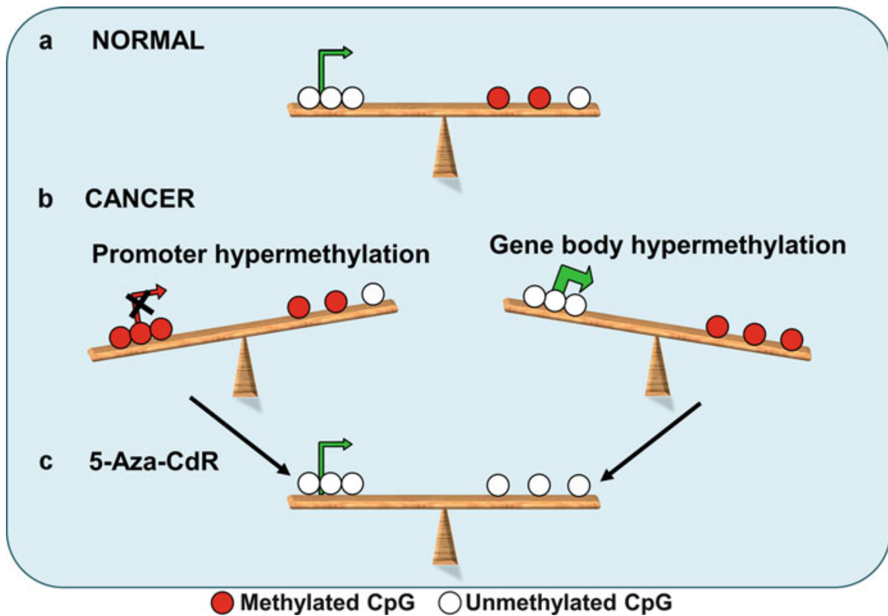


Fig. 1 DNA methylation equilibrium between the promoter and gene body modulates gene expression. In this diagram, methylated CpG sites are represented by *red circles*, unmethylated CpG sites are represented by *white circles*, and *green arrows* are indicative of active expression, while *red arrow* marks the absence of expression. (**a**) In normal mammalian tissue, genes that are actively transcribed have unmethylated promoters and some methylation in the gene body. (**b**) With the onset of cancer, however, promoter hypermethylation can turn off the expression of genes, and gene body hypermethylation can permit a more robust expression of some genes. (**c**) Treatment with DNA methyltransferase inhibitors such as 5-Aza-CdR can restore gene expression by removing aberrant methylation

the silencing of many known tumor suppressor genes (Fig. 1b) (Jones and Baylin 2007; Irizarry et al. 2009; Ehrlich and Lacey 2013; Shen and Laird 2013). Silencing of cell cycle regulators and DNA repair genes through DNA methylation has been reported in many different cancer types and is often mutually exclusive with the genetic inactivation of the gene (Sakai et al. 1991; Costello et al. 1996; Alvarez-Nuñez et al. 2006; Chiang et al. 2006). Sporadic breast and ovarian cancer display a loss of *BRCA1* expression due to promoter hypermethylation. Similarly, epigenetic silencing of tumor suppressor *VHL* via promoter methylation predisposes individuals to several malignancies including clear cell renal cell carcinoma (Herman et al. 1994; Esteller et al. 2000; Esteller 2001; Chiang et al. 2006; Creighton et al. 2013).

Silencing of DNA repair genes contributes to a greater burden of genomic instability and genetic mutations. O6-Methylguanine-DNA methyltransferase (*MGMT*), a DNA repair enzyme responsible for clearing out alkylation adducts on DNA, is frequently hypermethylated in many cancer types including gliomas and colorectal cancer. Consequently, *MGMT* was one of the first cancer DNA methylation biomarkers to be discovered. The suppression of *MGMT* due to promoter hypermethylation results in increased susceptibility to genetic mutations in essential genes such

as *p53* and *KRAS*. Interestingly, loss of MGMT makes the cell more vulnerable to treatment by chemotherapeutic agent temozolomide (TMZ). Clinical studies in glioblastoma multiforme (GBM) suggest that treatment with TMZ is most beneficial in cases where the tumor presents *MGMT* promoter hypermethylation (Donson et al. 2007; Silber et al. 2012; Zarnett et al. 2015).

Similarly, promoter hypermethylation of the mismatch repair gene *MLH1* is frequent in cancers. Studies have confirmed that the hypermethylation leads to an increased promoter nucleosome occupancy and decreased expression of *MLH1* (Lin et al. 2007). *MLH1* inactivation due to promoter methylation is strongly associated with hypermethylation of a subset of CpG islands, and it is the primary mechanism for microsatellite instability, contributing to the pathogenesis of many cancers including colorectal and endometrial carcinomas (Weisenberger et al. 2006; Hitchins et al. 2007; Hinoue et al. 2012; Li et al. 2013).

Hypermethylation of CpG islands can also contribute to the loss of imprinting. When the imprinted locus *IGF2/H19* becomes aberrantly methylated, the expression of the growth factor IGF2 is increased (Ravenel et al. 2001; Kaneda and Feinberg 2005). Sustained overexpression of IGF2 has been noted to contribute to the development and progression of cancers such as colorectal and gastric, and the loss of imprinting at this locus is the most common alteration in Wilms' tumor (Li et al. 1993; Taniguchi et al. 1995; Wu et al. 1997; Cui 2007; Bjornsson et al. 2007).

Aberrant DNA methylation is a widespread phenotype in cancer, and identifying the specific alterations driving the tumor phenotype can guide therapeutic strategies. In a recent study, our group applied the concept of DNA methylation addiction to identify epigenetic drivers of tumorigenesis. We hypothesized that cancer cells depend on the methylation of a few vital regions for survival, and these regions would be more likely to maintain DNA methylation when methylation levels were reduced artificially. Because these regions contribute to the fitness of the cancer cell, they are likely to be driving the tumor condition. To test this hypothesis, global DNA methylation of colorectal line HCT116 was compared to its lowly methylated derivative line lacking one or more of DNMTs (Rhee et al. 2000, 2002). Epigenetic drivers were ascertained by recognizing genomic regions that maintain methylation preferentially and in a cancer specific manner in the HCT116 derivative line. One of the candidate epigenetic drivers identified by this approach was interleukin-1 receptor-associated kinase 3 (*IRAK3)*. The *IRAK3* promoter is specifically hypermethylated in cancers, and this correlates with the reduced expression of the gene in many cancers including colon adenocarcinoma relative to normal tissue. Importantly, IRAK3 indirectly inhibits multiple pathways essential for cancer survival, including the STAT3, NF-kB, and MAPK pathways. Therefore, downregulation of *IRAK3* is greatly beneficial for cancer progression. Knocking down *IRAK3* in a non-tumorigenic cell line was sufficient to increase colony formation in vitro. *IRAK3* is silenced in HCT116 by DNA methylation, and overexpression of *IRAK3* in this line accounted for a decreased cell viability (De Carvalho et al. 2012).

5.1.2 Noncoding RNAs

Aside from canonical gene promoters, methylation also plays an important role in the regulation of noncoding RNA (ncRNA), such as microRNA (miRNA), small nucleolar RNA (snoRNA), vault RNA (vtRNA), and long noncoding RNA (lncRNA). These elements are critical regulators of cellular processes including proliferation, differentiation, and development (Esteller 2011). Aberrant hypermethylation can result in deregulation of microRNAs and contribute to cancer development. In bladder cancer cells, treatment with the DNMTi 5-Aza-2'-deoxycytidine (5-Aza-CdR) leads to the upregulation of miR-127 and the subsequent downregulation of the proto-oncogene BCL-6 (Saito et al. 2006; Ehrlich 2010; Kulis et al. 2013). Likewise, when the microRNA miR-124a becomes silenced due to hypermethylation in acute lymphoid leukemia (ALL), it activates the CDK6-RB1 oncogene pathway, contributing to poor patient survival (Agirre et al. 2009). It has also been observed in ALL that the CpG islands upstream of snoRNAs SNORD123, U70C, and ACA59B endure a cancer-specific hypermethylation resulting in their transcriptional silencing (Ferreira et al. 2012). Gastric cancer and acute myeloid leukemia (AML) patients with CpG hypermethylation of the ncRNA nc866, also known as vtRNA2-1, show poor survival (Treppendahl et al. 2012; Lee et al. 2014). *In vitro* knockdown of nc866 in gastric cell lines leads to the induction of known oncogenes, and overexpression of the ncRNA reduces cellular proliferation (Lee et al. 2014). In myelodysplastic syndrome (MDS), both vtRNA1-2 and vtRNA1-3 can be silenced by promoter methylation, and the hypermethylation of the vtRNA1-3 promoter is associated with a decreased survival in lower-risk MDS patients (Helbo et al. 2015). Finally, a recent study has detected epigenetic silencing of a partially annotated lncRNA *MORT* via DNA hypermethylation to be highly significant for the immortalization of human mammary epithelial cells. Deficient *MORT* expression is also common in most cancers and can be reactivated by 5-Aza-CdR treatment, suggesting a role for this lncRNA in immortalization during oncogenesis (Vrba et al. 2015). These findings and many others make it clear that aberrant methylation of ncRNAs with tumor suppression effects is a fundamental feature of cancer and has a vital role in the disease progression.

5.2 Hypomethylation

Although CpG hypomethylation was the first methylation change discovered in cancer, the implication of this dysregulation in tumorigenesis has often been overlooked. Feinberg and Vogelstein, as well as Gama-sosa et al., identified a global decrease in 5mC content across numerous cancer types (Gama-sosa et al. 1983; Feinberg and Vogelstein 1983). Hypomethylation can be an early event in tumorigenesis and is frequently detected in benign hyperplasia. Loss of methylation is more prominent with tumor progression, and metastatic lesions possess greater demethylation than primary tumors (Li et al. 2014b).

The majority of the decrease in CpG methylation occurs in intergenic and intragenic regions. These genomic areas are replete with repetitive and transposable elements. DNA methylation suppresses these elements and their hypomethylation can contribute to ectopic gene expression. Long interspersed nuclear element 1 (*LINE1*) retrotransposons are mobile genetic elements responsible for much of the endogenous mutagenesis in humans. *LINE1* insertions can greatly affect gene expression and DNA methylation is key to the silencing of *LINE1*. The hypomethylation of the CpG island at the promoter of *LINE1* stimulates the adoption of a permissive chromatin architecture at the alternative *MET* promoter, thereby activating the oncogene (Wolff et al. 2010). *LINE1* hypomethylation has also been recognized as an indicator of tumor progression and prognosis in several cancer types including prostate, melanoma, bladder, and renal cancer (Yegnasubramanian et al. 2008; Ecsedi et al. 2013; Andreotti et al. 2014; Su et al. 2014; Karami et al. 2015). Another class of repeat elements known as short interspersed nuclear elements (SINEs) is also similarly regulated by methylation, and studies have observed loss of methylation at these repeats in acute myeloid leukemia (AML) (Saied et al. 2012).

Although hypomethylation of non-CGI promoters is much less frequent than hypermethylation of promoter CGIs, it can result in the upregulation of oncogenes and proto-oncogenes (Feinberg and Vogelstein 1983; Søes et al. 2014). In metastatic non-small cell lung cancer tumors, for example, the putative oncogene engulfment and cell motility 3 (*ELMO3*) gene is significantly overexpressed as a result of its promoter hypomethylation (Søes et al. 2014). In osteosarcoma, Iroquois homeobox 1 (*IRX1)* is upregulated and is pro-metastatic. The increase of *IRX1* gene expression is found in both metastatic osteosarcoma cell lines and primary patient samples (Lu et al. 2015). In both cases the gain in expression is associated with hypomethylation of the gene promoter.

5.3 DNA Methylation at Intergenic and Intragenic Regions in Cancer

For decades, much of the research efforts in cancer epigenetics had been concentrated on the regulation of DNA methylation at gene promoters. Advances in next-generation and high-density array sequencing have allowed researchers to expand their studies of DNA methylation to a genome-wide context. In doing so, it has become increasingly evident that non-promoter intragenic and intergenic regions are also dynamically regulated and contribute to physiological changes as well as to the development of disease states.

5.3.1 DNA Methylation Changes in Transcribed Regions

Unlike promoters, where methylation contributes to a “closed” chromatin architecture resulting in gene repression, the methylation level in transcribed regions (bodies) of genes is often positively correlated with gene expression. A recent investigation of glioblastoma samples revealed functional roles for gene body methylation in affecting *MGMT* expression (Moen et al. 2014). The study found that tumors with unmethylated *MGMT* promoter and high gene body methylation maintained

a high MGMT expression. As previously mentioned, MGMT expression confers resistance to TMZ therapy. Consequently, pretreating glioblastoma cell lines with DNMTi decitabine to reduce MGMT body methylation significantly sensitized them to the temozolomide treatment (Moen et al. 2014).

Gene bodies are mostly CpG poor, contain numerous repetitive and transposable elements, and are extensively methylated. While DNA methylation inhibits initiation of transcription, it enables transcription elongation (Kulis et al. 2013; Lou et al. 2014). Furthermore, methylation in the gene body can also add to transcription efficiency by regulating the usage of alternate start sites. Global methylome analysis of GBMs purports a role for gene body hypomethylation in stimulating the transcription from alternate promoters resulting in an increased expression of alternative transcripts and expression of oncogenic protein isoforms (Nagarajan et al. 2014). Finally, loss of methylation in gene bodies can reveal distal regulatory elements (enhancers) that might have been muted tissue specifically. A recent large-scale analysis comparing DNA methylation profiles of normal B cell and chronic myeloid leukemia revealed widespread gene body hypomethylation targeting particularly enhancer sites (Kulis et al. 2012).

5.3.2 DNA Methylation and Enhancers

Along with promoters, enhancers play a significant role in regulating the expression and activity of target genes. Enhancers serve as a platform for transcription factors (TFs), which bind the DNA through sequence recognition. The presence of multiple TFs at the enhancer is usually necessary for enhancer activation. Additionally, functional enhancers are decorated with active histone marks including H3K4me1 and H3K27ac. Through long-range interactions such as "looping," these distal elements are able to deliver the bound accessory proteins to promoters and stimulate robust transcription. Of note, each enhancer can regulate the activity of multiple promoters (Bulger and Groudine 2011).

Although methylation of DNA has been noted to be inversely correlated with the presence of active histone marks, such as those that delineate active enhancers (Lay et al. 2015; Jones 2012; Kelly et al. 2012), expression-related methylation sites co-localizing with enhancers have also been observed. Not only is methylation at these sites inversely correlated with gene expression, similar to promoters, but they are often better predictors of expression levels than the promoter methylation (Aran et al. 2013; Aran and Hellman 2013). Furthermore, enhancers can regulate gene expression in a cell-type-specific manner even when the promoter is continually unmethylated (Aran et al. 2013).

TF recognition sequences and other DNA-binding elements are mostly situated in unmethylated DNA. DNA methylation can thwart the association of TF to DNA, and conversely, the presence of TFs can promote DNA hypomethylation by preventing DNMTs from accessing DNA (Calo and Wysocka 2013). Thus, subtle modulation of DNA methylation at enhancers can greatly affect gene expression of multiple target genes.

In cancer, hypomethylation of intergenic and intragenic enhancers can reveal binding motifs for TFs and induce downstream expression changes (Kulis et al. 2013; Aran et al. 2013). On the other hand, DNA hypermethylation at enhancers

can decommission them, resulting in a loss of active histone marks and loss of transcription factor binding. Such alterations can modulate gene transcription independent of promoter methylation fluctuations (Kulis et al. 2013).

6 Tumor Stratification and DNA Methylation Marker Discovery Accelerated by TCGA

6.1 Consortium Data

Recent technological advancements in DNA sequencing have made it feasible to generate genome-wide genetic and epigenetic profiles for numerous tumor and normal samples. Integrating all the various datasets allows us to construct a more complete picture of how the different constituents of the tumor machinery contribute to the initiation and progression of malignant tumors. Moreover, large sample sizes and available patient information make it feasible to stratify tumors into subgroups that can be tackled as unique entities for more personalized and effective treatment options. However, numerous bioinformatics and logistic challenges arise with such large datasets. To address these challenges, many research groups have come together to work in multinational consortia, such as Encyclopedia of DNA Elements (ENCODE), NIH Roadmap Epigenomics Mapping Consortium, and The Cancer Genome Atlas (TCGA). The ENCODE project has surveyed a number of cell lines to extrapolate functional and regulatory elements of the genome, and the NIH Roadmap has focused its resources on interrogating various tissue types to identify tissue-specific regulation of the epigenome, while TCGA has comprehensively collected data from 10,000 tumor samples across 30 cancer types (ENCODE Project Consortium 2004; Birney et al. 2007; Bernstein et al. 2010; Chadwick 2012; ENCODE Project Consortium 2012; Weinstein et al. 2013; Tomczak et al. 2015). Molecular profiles generated by TCGA include whole-exome sequencing for mutational information, RNA sequencing of the transcriptome, single-nucleotide polymorphism (SNP) arrays to determine somatic copy number variations, and Illumina Infinium Bead Array analysis of global methylation status (Tomczak et al. 2015). Along with the molecular information, TCGA also gathers details on tumor grade, stage, and prognosis. Researches, therefore, have been able to take advantage of the vast treasure trove of molecular data generated by these consortia to stratify cancer types into subgroups, gain insights into the mechanisms specific to these subgroups, and identify subgroup-specific therapeutic targets (Weisenberger 2014).

6.2 CpG Island Methylator Phenotype (CIMP) Stratifies Tumor Subclass

In 1999, Toyota et al. noted that a subset of colorectal cancers showed cancer-specific hypermethylation of specific CpGs. Moreover, this subset of tumors displayed a concordant hypermethylation of *p16*, *THBS1*, and *hMLH1* promoters. The group coined this phenomenon as CpG island methylator phenotype (CIMP). They further

postulated that CIMP contributes to tumorigenesis by concurrently incapacitating multiple tumor suppressor genes through hypermethylation of their respective CGI promoters (Toyota et al. 1999). In 2006, Weisenberger and colleagues utilized methylation data from CRC samples to identify a panel of markers that identified the CIMP-positive tumors. This subset of CRC tumors robustly correlated with the ^{v600E}*BRAF* mutation and microsatellite instability (Weisenberger et al. 2006). While the molecular basis for the onset of CIMP in CRC is still unclear, several studies have now unequivocally proven its existence.

One of the first mechanistic insights into CIMP generation in cancer came from investigating promoter-associated hypermethylation in gliomas. Using TCGA data, Noushmehr et al. comprehensively characterized DNA methylation of GBM tumor and identified a CIMP type that defines a subset of gliomas. Interestingly, G-CIMP tumors were tightly associated with a high frequency of *isocitrate dehydrogenase 1* (*IDH1*) somatic mutations (Noushmehr et al. 2010; Brennan et al. 2013). Somatic mutations of *IDH1* confer gain of function activity in the mutant isoform allowing the mutated protein to produce 2-hydroxygluterate (2-HG). This oncometabolite is an inhibitor of the TET family dioxygenases and Jumonji-C domain containing histone lysine demethylases. Thus, production of 2-HG results in the accumulation of DNA methylation along with aberrant histone methylation (Dang et al. 2009). More recent studies have shown that the IDH1 mutation alone is sufficient to establish a hypermethylator phenotype in gliomas and that this hypermethylator status is retained in both early and late tumor of the same patient, suggesting that CIMP phenotype is an early event that is likely driving the tumorigenesis (Turcan et al. 2012; Hill et al. 2014).

Similar to GBMs, AML tumors bear *IDH1*and *IDH2* mutations as well as *TET* mutations. *IDH1* and *IDH2* mutations are mutually exclusive, while *TET2* mutations are mutually exclusive with all *IDH* mutations, suggesting redundant activity of the proteins. TCGA and others have shown that AML tumors with mutations in IDH proteins or TET enzymes show substantial DNA hypermethylation (Figueroa et al. 2010; Shih et al. 2012).

To date, several reports have described CIMPs in many additional cancers including gastric, breast, bladder, melanoma, prostate, hepatocellular, and endometrial cancer. Stratifying cancers into subsets according to DNA methylation can provide valuable prognostic, diagnostic, and therapeutic insights. In the case of GBMs, G-CIMP patients tend to be younger in age and have better survival outcomes than the non-G-CIMP patients. Similarly, Fang and colleagues found that B-CIMP+ breast tumors were associated with estrogen receptor (ESR1)-/progesterone receptor (PGR)-positive tumors, and the CIMP status was a strong prognosis indicator. B-CIMP+ patients had a lower risk of metastasis and better clinical survival (Fang et al. 2011). Recognizing and understanding the onset of the methylator phenotype can thus help researchers to better strategize therapeutic options.

6.3 DNA Methylation-Based Biomarkers

DNA methylation is an extremely stable mark, and the methylation status of loci can be readily obtained from blood, urine sediments, and even highly processed

tissues. Thus, markers based on the DNA methylation status of CpG sites are convenient prognostic and diagnostic tools (Laird 2003; Levenson 2010).

Being able to integrate DNA methylation data with gene expression profiles of hundreds of tumor and normal samples permits the discovery of individual tumor suppressors and oncogenes, as well as the identification of methylation signatures such as CIMP. In addition, this vast data trove can be mined for biomarker identification and validation. Using TCGA high-grade serous ovarian cancer datasets, researchers have been able to identify promoter methylation events in 168 genes, including *BRCA1*. Inactivation of *BRCA1* due to promoter methylation and mutations of the locus are mutually exclusive. While high-grade serous ovarian patients carrying genetic mutations in *BRCA1* show better overall survival than patients with *BRCA1* wild-type gene, interestingly, patients with epigenetic silencing of *BRCA1* do not carry this survival advantage (TCGA 2011).

Comprehensive examination of 446 clear cell renal cell carcinomas (ccRCC) led to the recognition of *UQCRH* as a putative tumor suppressor in ccRCC. Hypermethylation of the locus was observed in 36% of the tumors and it correlated with higher stage and grade. Additionally, by correlating clinical outcomes with protein signatures, it became evident that a glycolytic shift similar to the "Warburg effect" occurs in ccRCC. One of the drivers of this shift was the promoter hypomethylation of *MIR21*, a negative regulator of the tumor suppressor PTEN. The loss of promoter methylation correlated with increased expression of *MIR21* and was associated with a worse patient outcome (Creighton et al. 2013).

Researchers outside the consortium have also been able to use repository to discover and validate biomarkers. For example, using methylation data of 194 AML patients collected by TCGA, a recent study identified a CpG site in the complement component 1 subcomponent R (C1R) to be a strong predictor of overall survival. Patients with high levels of cytosine methylation at this site showed a significantly longer overall survival than those with low levels of methylation (Božić et al. 2015).

7 DNA Methylation as a Therapeutic Target

Epigenetic aberrations in cancers including differential DNA methylation can be used to distinguish tumor subtypes, indicate treatment responsiveness, predict clinical outcomes, and determine therapeutic strategies. Epigenetic profiles can reveal molecular pathways most vulnerable to chemotherapeutic agents, and methylation changes can often serve as a barometer for treatment efficacy (Kelly et al. 2010). Unlike genetic modifications, DNA methylation is both somatically heritable and reversible. Thus, DNA methylation changes affected through pharmacological intervention can have long-lasting impact. In addition, cancer cells can become addicted to the advantages rendered by the atypical methylation landscape making them increasingly vulnerable to epigenetic therapy (Mair et al. 2014). To this end, DNMT inhibitors have been successfully employed in preclinical and clinical settings with the goal of eliminating aberrant methylation (Yamazaki and Issa 2013; Juo et al. 2015).

DNA methyltransferase inhibitors (DNMTi), such as the cytidine analogs 5-Aza-2'-deoxycytidine (5-Aza-CdR) and decitabine, become incorporated into DNA during replication and are recognized as natural substrate by DNMTs. The DNMT initiates the methylation reaction by covalently binding DNA. The resolution of this covalent bond is impeded by Aza-cytosine, and the covalent sequestering of DNMTs to DNA concedes the integrity of the DNA molecule and elicits DNA damage response. This triggers proteomic degradation of the bound DNMT contributing to the subsequent loss of methylation marks (Christman 2002; Stresemann and Lyko 2008).

5-Azacytidine is currently FDA approved to treat high-risk MDS patients and has resulted in successful clinical outcomes (Fenaux et al. 2009). Preclinical data are also available for other cytidine analogs, such as S110 which shows better stability and activity relative to 5-Aza-CdR (Yoo et al. 2007; Chuang et al. 2010). Treatment by DNMTi can sensitize cancers to other chemotherapeutic agents such as in the case of administrating SGI-110 to hepatocellular carcinoma cells. SGI-110 significantly synergized with oxaliplatin and resulted in greater cytotoxicity (Kuang et al. 2015).

DNMTi can also prove to be immunomodulatory. The hypomethylation induced in epithelial ovarian carcinoma cells upon treatment with SGI-110 results in the increased expression of cancer-testis antigens, thereby, enhancing the recognition of EOC cells by antigen-specific CD8+ T-cells. This contributes to restricted tumor growth and better survival in a xenograft setting (Srivastava et al. 2015).

Additionally, numerous tumor suppressor gene promoter targets of DNMTi have been identified, including *p16*, *MYOD1*, *RASSF1A*, and *TIMP3* (Toyota 2001; Christman 2002). The chromatin remodeler protein CHD5 is considered a tumor suppressor in many cancer types and is frequently silenced through multiple epigenetic mechanisms including promoter hypermethylation (Fujita et al. 2008; Gorringe et al. 2008; Wang et al. 2011). A study performed in a colorectal cancer model found that treatment with 5-Aza-CdR partially restored CHD5 protein expression (Fatemi et al. 2014). In AML cells, researchers have found that DNMTi can initiate apoptosis in a p53-independent manner. Here, 5-Aza-CdR administration can demethylate the promoter of *p73*, a member of the p53 family of transcription factors. The expression of TP73 induces p21 protein expression, which in turn renders the cell more sensitive to chemotherapeutics and mediates the cytotoxicity of the drug (Schmelz et al. 2005).

Better understanding of the role of intergenic methylation in the recent years has led researchers to realize that in addition to promoter methylation, gene body methylation might also serve as a therapeutic target for demethylating chemotherapeutic agents such as 5-Aza-CdR (Fig. 1b, c). In our recent study, genome-wide methylation levels were assayed at various time points after a short treatment of the colorectal cell line HCT116 with 5-Aza-CdR. The study not only confirmed that loss of methylation from the gene body correlated with loss of gene expression, but importantly, the rate of re-methylation after drug withdrawal determined the strength of reexpression. By taking advantage of HCT116 derivative lines lacking various DNA methyltransferases, the study was able to conclude that the re-methylation was dependent on DNMT3B. Moreover, clustering the genomic regions into

groups according to the rates of re-methylation, researchers noticed that rapidly re-methylating genes are enriched for oncogenic genes such as c-MYC targets and metabolic pathway genes. Thus, a potential mechanism of action for DNA methyltransferase inhibitors could be through mitigating the effect of deregulated c-MYC (Kasinathan and Henikoff 2014; Yang et al. 2014). This study defined a causal link between gene body DNA methylation and gene expression and recognized that gene body methylation can be targeted to lower gene expression of oncogenes critical to cancer progression.

The effects of DNMT inhibitors are diverse, and therapeutic responses have a slow onset. Additionally, low doses of DNMTi are sufficient for long-lasting loss of tumorigenicity and self-renewal with minimal cytotoxic effect. All of this indicates that supplementary to acute reexpression of tumor suppressor genes or downregulation of crucial oncogene, other mechanism(s) must exist by which DNA methyltransferase inhibitors can target methylation (Oki et al. 2007; Tsai et al. 2012; Licht 2015). Recent investigations, including a study from our group, have shown that the demethylating agents might be mediating therapeutic response by rendering the cell more visible to the immune system (Chiappinelli et al. 2015; Roulois et al. 2015). Specifically, demethylating agents are able to trigger the induction of an antiviral immune response by permitting the expression of endogenous retroviruses that had previously been silenced by DNA methylation (Fig. 2).

Stimulation of immune response has been long recognized as a function of DNMTi. Early studies have noted that 5-Aza-CdR can demethylate and activate tumor antigens (De Smet et al. 1996; Shiohama et al. 2014). In non-small cell lung cancer (NSCLC), it has been observed that upon treatment of 5-Aza-CdR, some of the patients have a robust response to immune checkpoint blockade therapy, suggesting that the DNMTi might have sensitized this cohort to the immune checkpoint inhibition (Wrangle et al. 2013). In NSCLC and others, treatment with Aza has been shown to stimulate a strong upregulation of interferon pathway genes along with increased expression of endogenous retroviral (ERV) transcripts. Moreover, interferon genes and genes involved in antigen presentation accounted for the majority of genes commonly upregulated in solid tumor cell lines upon Aza treatment (Li et al. 2014a).

To better understand the underlying mechanism of DNMTi, a study from our group following up on our previous work (Yang et al. 2010) focused on the effect of transient low-dose 5-Aza-CdR treatment of colorectal cell lines. Through gene expression profiling, the study determined that the majority of the late occurring expression changes (24 days past initial exposure) were of interferon-responsive genes. These genes showed little modulation of methylation at their promoters or coding region, and in fact, many of them displayed low DNA methylation levels pretreatment. Thus, it can be interpreted that the change in gene expression upon treatment with 5-Aza-CdR is independent of the drug's capacity to demethylate the respective gene promoters. A series of genetic experiments provided sufficient evidence to the claim that the activation of these genes occurred through the RIG1/MDA5/MAVS/IRF7 signaling pathway. RIG1 and MDA5 are cytosolic pattern recognition receptors whose primary role is to recognize viral RNA (RIG1

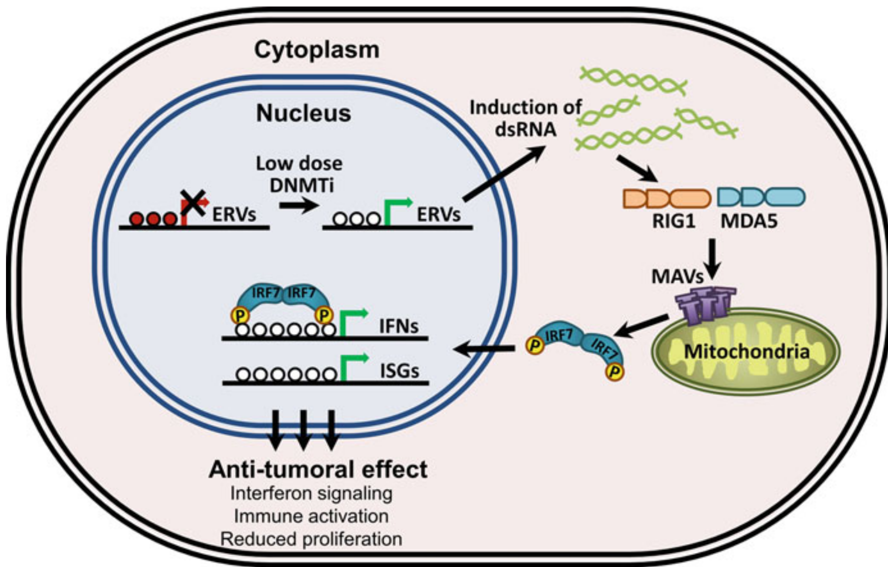


Fig. 2 DNMTi exert antitumoral effect by eliciting immune response in cancer cells. Treatment with DNA methyltransferase inhibitors induces transcription of endogenous retroviral (ERV) elements. These double-stranded RNAs are recognized by viral recognition proteins such as RIG1 and MDA5, which in turn interact with the mitochondrial antiviral signaling (MAVS) proteins. MAVS-mediated IRF7 activation leads to the translocation of IRF7 from the cytoplasm to the nucleus where it initiates transcription of interferons (IFNs) and interferon-stimulated genes (ISGs), which then contribute to reduced proliferation (Modified from Chiappinelli et al. 2015; Licht 2015; Roulois et al. 2015)

recognizes single- and double-stranded RNA (dsRNA), while MDA5 recognizes double-stranded RNA) and initiate a signaling cascade dependent on the mitochondrial antiviral signaling (MAVS) adaptor molecule. This leads to the activation of downstream targets, such as IRF7, and culminates in a strong antitumor response. The study found that 5-Aza-CdR induced a significant increase of dsRNAs including a robust induction of endogenous retrovirus RNA transcription (Roulois et al. 2015). Another group working with ovarian cancer cell lines came to a similar conclusion that treatment with 5-Aza-CdR triggered the upregulation of interferon signaling mediated by downstream activity of IRF7. Furthermore, the strength of interferon response to the drug treatment was reflective of how well the tumor would respond to the immune checkpoint therapy (Chiappinelli et al. 2015). Thus, a major mode of action of DNMT inhibitors such as 5-Aza-CdR is the loss of DNA methylation at previously silenced repetitive elements, such as ERVs, and the subsequent induction of dsRNA transcription triggers a strong antiviral response. As a consequence, there is an overall antitumoral effect including interferon induction, reduced cell proliferation, and loss of self-renewal capacity upon treatment (Fig. 2). Furthermore, the incorporation of DNMTi will be dependent on cell doubling time (Bender et al. 1999). Cancer cells tend to have shorter doubling times and higher

rates of metabolism than normal cells (Cheng et al. 2004). Thus, the cancer cells will be more affected by the treatment and show a stronger production of dsRNA and subsequent immune response.

8 Concluding Remarks

DNA methylation is a complex epigenetic mechanism crucial to regulating gene expression in normal and tumor cells. Methylation of CpGs at the promoters of genes attenuates their expression, while gene body methylation levels positively correlate with expression. By modulating gene expression, DNA methylation is able to alter signaling pathways that affect cellular processes such as cell cycle, DNA repair, cell growth, and proliferation. Dysregulation of DNA methylation can, therefore, lead to inappropriate silencing of tumor suppressors or expression of oncogenes, thus contributing to the development of disease states including cancer. However, unlike genetic changes, DNA methylation alterations can be potentially reversed with the help of methylation inhibitors. This can achieve therapeutic effects by reactivating silenced tumor suppressor genes, downregulating overexpressed oncogenes, and stimulating immune response toward cancer cells. Genome-wide screens can be efficiently used to identify genes that are influenced by the pathways being affected by aberrant methylation. Furthermore, with improved access to next-generation sequencing, large-scale multinational consortia led research that has resulted in a wealth of genomic and epigenomic data. Integrating this information with patient profiles will enable researchers to validate putative therapeutic epigenetic targets, as well as stratify tumors into clinically relevant subgroups according to their methylation status, thereby, allowing to design more effective therapeutic strategies.

Acknowledgment The work in the Liang laboratory has been supported in part by the generous contribution of George and Vicky Joseph.

References

Agirre X, Vilas-Zornoza A, Jiménez-Velasco A, et al. Epigenetic silencing of the tumor suppressor microRNA Hsa-miR-124a regulates CDK6 expression and confers a poor prognosis in acute lymphoblastic leukemia. Cancer Res. 2009;69:4443–53. doi:10.1158/0008-5472.CAN-08-4025.

Alvarez-Nuñez F, Bussaglia E, Mauricio D, et al. PTEN promoter methylation in sporadic thyroid carcinomas. Thyroid. 2006;16:17–23. doi:10.1089/thy.2006.16.17.

Andreotti G, Karami S, Pfeiffer RM, et al. LINE1 methylation levels associated with increased bladder cancer risk in pre-diagnostic blood DNA among US (PLCO) and European (ATBC) cohort study participants. Epigenetics. 2014;9:404–15. doi:10.4161/epi.27386.

Aran D, Hellman A. DNA methylation of transcriptional enhancers and cancer predisposition. Cell. 2013;154:11–3. doi:10.1016/j.cell.2013.06.018.

Aran D, Sabato S, Hellman A. DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. Genome Biol. 2013;14:R21. doi:10.1186/gb-2013-14-3-r21.

Baylin SB, Jones PA. A decade of exploring the cancer epigenome — biological and translational implications. Nucleus. 2011. doi:10.1038/nrc3130.

Bender CM, Gonzalgo ML, Gonzales FA, et al. Roles of cell division and gene transcription in the methylation of CpG islands. Mol Cell Biol. 1999;19:6690–8.

Bernstein BE, Stamatoyannopoulos JA, Costello JF, et al. The NIH roadmap epigenomics mapping consortium. Nat Biotechnol. 2010;28:1045–8. doi:10.1038/nbt1010-1045.

Bestor TH, Edwards JR, Boulard M. Notes on the role of dynamic DNA methylation in mammalian development. Proc Natl Acad Sci. 2015;112:6796–9. doi:10.1073/pnas.1415301111.

Birney E, Stamatoyannopoulos JA, Dutta A, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007;447:799–816. doi:10.1038/nature05874.

Bjornsson HT, Brown LJ, Fallin MD, et al. Epigenetic specificity of loss of imprinting of the IGF2 gene in Wilms tumors. J Natl Cancer Inst. 2007;99:1270–3. doi:10.1093/jnci/djm069.

Božić T, Lin Q, Frobel J, et al. DNA-methylation in C1R is a prognostic biomarker for acute myeloid leukemia. Clini Epigenetics. 2015;7:116. doi:10.1186/s13148-015-0153-6.

Brennan CW, Verhaak RGW, McKenna A, et al. The somatic genomic landscape of glioblastoma. Cell. 2013;155:462–77. doi:10.1016/j.cell.2013.09.034.

Bulger M, Groudine M. Functional and mechanistic diversity of distal transcription enhancers. Cell. 2011;144:327–39. doi:10.1016/j.cell.2011.01.024.

Calo E, Wysocka J. Modification of enhancer chromatin: what, how, and why? Mol Cell. 2013;49:825–37. doi:10.1016/j.molcel.2013.01.038.

Chadwick LH. The NIH roadmap epigenomics program data resource. Epigenomics. 2012;4:317–24. doi:10.2217/epi.12.18.

Cheng JC, Yoo CB, Weisenberger DJ, et al. Preferential response of cancer cells to zebularine. Cancer Cell. 2004;6:151–8. doi:10.1016/j.ccr.2004.06.023.

Chiang JW, Karlan BY, Cass L, Baldwin RL. BRCA1 promoter methylation predicts adverse ovarian cancer prognosis. Gynecol Oncol. 2006;101:403–10. doi:10.1016/j.ygyno.2005.10.034.

Chiappinelli KB, Strissel PL, Desrichard A, et al. Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. Cell. 2015;162:974–86. doi:10.1016/j.cell.2015.07.011.

Christman JK. 5-Azacytidine and 5-aza-2’-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. Oncogene. 2002;21:5483–95. doi:10.1038/sj.onc.1205699.

Chuang JC, Warner SL, Vollmer D, et al. S110, a 5-Aza-2’-deoxycytidine-containing dinucleotide, is an effective DNA methylation inhibitor in vivo and can reduce tumor growth. Mol Cancer Ther. 2010;9:1443–50. doi:10.1158/1535-7163.MCT-09-1048.

Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. Nat Rev Genet. 2009;10:691–703. doi:10.1038/nrg2640.

Costello JF, Berger MS, Huang HS, Cavenee WK. Silencing of p16/CDKN2 expression in human gliomas by methylation and chromatin condensation. Cancer Res. 1996;56:2405–10.

Creighton CJ, Morgan M, Gunaratne PH, et al. Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature. 2013;499:43–9. doi:10.1038/nature12222.

Cui H. Loss of imprinting of IGF2 as an epigenetic marker for the risk of human cancer. Dis Markers. 2007;23:105–12.

Dang L, White DW, Gross S, et al. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. Nature. 2009;462:739–44. doi:10.1038/nature08617.

De Carvalho DD, Sharma S, You JS, et al. DNA methylation screening identifies driver epigenetic events of cancer cell survival. Cancer Cell. 2012;21:655–67. doi:10.1016/j.ccr.2012.03.045.

De Smet C, De Backer O, Faraoni I, et al. The activation of human gene MAGE-1 in tumor cells is correlated with genome-wide demethylation. Proc Natl Acad Sci U S A. 1996;93:7149–53.

Donson AM, Addo-Yobo SO, Handler MH, et al. MGMT promoter methylation correlates with survival benefit and sensitivity to temozolomide in pediatric glioblastoma. Pediatr Blood Cancer. 2007;48:403–7. doi:10.1002/pbc.20803.

Eckhardt F, Lewin J, Cortese R, et al. DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet. 2006;38:1378–85. doi:10.1038/ng1909.

Ecsedi SI, Hernandez-Vargas H, Lima SC, et al. Transposable hypomethylation is associated with metastatic capacity of primary melanomas. Int J Clin Exp Pathol. 2013;6:2943–8.

Ehrlich M. DNA hypomethylation in cancer cells. Epigenomics. 2010;1:239–59. doi:10.2217/epi.09.33.DNA.

Ehrlich M, Lacey M. Epigenetic alterations in oncogenesis. Epigenetic Alterations Oncog Adv Exp 31 Med Biol. 2013;754:31–56. doi:10.1007/978-1-4419-9967-2.

ENCODE Project Consortium. The ENCODE (ENCyclopedia of DNA Elements) Project. Science. 2004;306:636–40. doi:10.1126/science.1105136.

ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489:57–74. doi:10.1038/nature11247.

Esteller M. DNA methylation patterns in hereditary human cancers mimic sporadic tumorigenesis. Hum Mol Genet. 2001;10:3001–7. doi:10.1093/hmg/10.26.3001.

Esteller M. Non-coding RNAs in human disease. Nat Rev Genet. 2011;12:861–74. doi:10.1038/nrg3074.

Esteller M, Silva JM, Dominguez G, et al. Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. J Natl Cancer Inst. 2000;92:564–9.

Fang F, Turcan S, Rimner A, et al. Breast cancer methylomes establish an epigenomic foundation for metastasis. Sci Transl Med. 2011;3:75ra25. doi:10.1126/scitranslmed.3001875.

Fatemi M, Paul TA, Brodeur GM, et al. Epigenetic silencing of CHD5, a novel tumor-suppressor gene, occurs in early colorectal cancer stages. Cancer. 2014;120:172–80. doi:10.1002/cncr.28316.

Feinberg AP, Vogelstein B. Hypomethylation of ras oncogenes in primary human cancers. Biochem Biophys Res Commun. 1983;111:47–54. doi:10.1016/S0006-291X(83)80115-6.

Fenaux P, Mufti GJ, Hellstrom-Lindberg E, et al. Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: a randomised, open-label, phase III study. Lancet Oncol. 2009;10:223–32. doi:10.1016/S1470-2045(09)70003-8.

Ferguson-Smith AC. Genomic imprinting: the emergence of an epigenetic paradigm. Nat Rev Genet. 2011;12:565–75. doi:10.1038/nrg3032.

Ferreira HJ, Heyn H, Moutinho C, Esteller M. CpG island hypermethylation-associated silencing of small nucleolar RNAs in human cancer. RNA Biol. 2012;9:881–90. doi:10.4161/rna.19353.

Figueroa ME, Abdel-Wahab O, Lu C, et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. Cancer Cell. 2010;18:553–67. doi:10.1016/j.ccr.2010.11.015.

Fujita T, Igarashi J, Okawa ER, et al. CHD5, a tumor suppressor gene deleted from 1p36.31 in neuroblastomas. J Natl Cancer Inst. 2008;100:940–9. doi:10.1093/jnci/djn176.

Gama-sosa MA, Slagel VA, Trewyn RW. The 5-methylcytosine content of DNA from human tumors. Nucleic Acids Res. 1983;11:6883–94. doi:10.1093/nar/11.19.6883.

Gardiner-Garden M, Frommer M. CpG islands in vertebrate genomes. J Mol Biol. 1987;196:261–82.

Gorringe KL, Choong DY, Williams LH, et al. Mutation and methylation analysis of the chromodomain-helicase-DNA binding 5 gene in ovarian cancer. Neoplasia. 2008;10:1253–8.

Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144:646–74.

Helbo A, Treppendahl M, Aslan D, et al. Hypermethylation of the VTRNA1-3 promoter is associated with poor outcome in lower risk myelodysplastic syndrome patients. Genes (Basel). 2015;6:977–90. doi:10.3390/genes6040977.

Herman JG, Latif F, Weng Y, et al. Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. Proc Natl Acad Sci U S A. 1994;91:9700–4.

Hill VK, Shinawi T, Ricketts CJ, et al. Stability of the CpG island methylator phenotype during glioma progression and identification of methylated loci in secondary glioblastomas. BMC Cancer. 2014;14:506. doi:10.1186/1471-2407-14-506.

Hinoue T, Weisenberger DJ, Lange CPE, et al. Genome-scale analysis of aberrant DNA methylation in colorectal cancer. Genome Res. 2012;22:271–82. doi:10.1101/gr.117523.110.

Hitchins MP, Wong JJL, Suthers G, et al. Inheritance of a cancer-associated MLH1 germ-line epimutation. N Engl J Med. 2007;356:697–705. doi:10.1056/NEJMoa064522.

Irizarry RA, Ladd-Acosta C, Wen B, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet. 2009;41:178–86. doi:10.1038/ng.298.
Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. Nat Rev Genet. 2002;3:415–28. doi:10.1038/nrg816.
Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012;13:484–92. doi:10.1038/nrg3230.
Jones PA, Baylin SB. The epigenomics of cancer. Cell. 2007;128:683–92. doi:10.1016/j.cell.2007.01.029.
Jones PA, Laird PW. Cancer epigenetics comes of age. Nat Genet. 1999;21:163–7. doi:10.1038/5947.
Jones PA, Liang G. Rethinking how DNA methylation patterns are maintained. Nat Rev Genet. 2009;10:805–11. doi:10.1038/nrg2651.
Juo Y-Y, Gong X-J, Mishra A, et al. Epigenetic therapy for solid tumors: from bench science to clinical trials. Epigenomics. 2015;7:215–35. doi:10.2217/epi.14.73.
Kaneda A, Feinberg AP. Loss of imprinting of IGF2: a common epigenetic modifier of intestinal tumor risk. Cancer Res. 2005;65:11236–40. doi:10.1158/0008-5472.CAN-05-2959.
Karami S, Andreotti G, Liao LM, et al. LINE1 methylation levels in pre-diagnostic leukocyte DNA and future renal cell carcinoma risk. Epigenetics. 2015;10:282–92. doi:10.1080/15592294.2015.1006505.
Kasinathan S, Henikoff S. 5-Aza-CdR delivers a gene body blow. Cancer Cell. 2014;26:449–51. doi:10.1016/j.ccell.2014.09.004.
Kelly TK, De Carvalho DD, Jones PA. Epigenetic modifications as therapeutic targets. Nat Biotechnol. 2010;28:1069–78. doi:10.1038/nbt.1678.
Kelly TK, Liu Y, Lay FD, et al. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. Genome Res. 2012;22:2497–506. doi:10.1101/gr.143008.112.
Ko M, Huang Y, Jankowska AM, et al. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature. 2010;468:839–43.
Kohli RM, Zhang Y. TET enzymes, TDG and the dynamics of DNA demethylation. Nature. 2013;502:472–9. doi:10.1038/nature12750.
Kuang Y, El-Khoueiry A, Taverna P, et al. Guadecitabine (SGI-110) priming sensitizes hepatocellular carcinoma cells to oxaliplatin. Mol Oncol. 2015;9:1799–814. doi:10.1016/j.molonc.2015.06.002.
Kulis M, Heath S, Bibikova M, et al. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. Nat Genet. 2012;44:1236–42. doi:10.1038/ng.2443.
Kulis M, Queirós AC, Beekman R, Martín-Subero JI. Intragenic DNA methylation in transcriptional regulation, normal differentiation and cancer. Biochim Biophys Acta Gene Regul Mech. 2013;1829:1161–74. doi:10.1016/j.bbagrm.2013.08.001.
Laird PW. The power and the promise of DNA methylation markers. Nat Rev Cancer. 2003;3:253–66. doi:10.1038/nrc1045.
Lay FD, Liu Y, Kelly TK, et al. The role of DNA methylation in directing the functional organization of the cancer epigenome. 2015:1–11. doi:10.1101/gr.183368.114.Freely.
Lee K-S, Park J-L, Lee K, et al. nc886, a non-coding RNA of anti-proliferative role, is suppressed by CpG DNA methylation in human gastric cancer. Oncotarget. 2014;5:3944–55.
Levenson VV. DNA methylation as a universal biomarker. Expert Rev Mol Diagn. 2010;10:481–8. doi:10.1586/erm.10.17.
Li E, Beard C, Jaenisch R. Role for DNA methylation in genomic imprinting. Nature. 1993;366:362–5. doi:10.1038/366362a0.
Li H, Chiappinelli KB, Guzzetta AA, et al. Immune regulation by low doses of the DNA methyltransferase inhibitor 5-azacitidine in common human epithelial cancers. Oncotarget. 2014a;5:587–98.
Li J, Huang Q, Zeng F, et al. The prognostic value of global DNA hypomethylation in cancer: a meta-analysis. PLoS One. 2014b;9:e106290. doi:10.1371/journal.pone.0106290.

Li X, Yao X, Wang Y, et al. MLH1 promoter methylation frequency in colorectal cancer patients and related clinicopathological and molecular features. PLoS One. 2013;8:e59064. doi:10.1371/journal.pone.0059064.

Liang G, Chan MF, Tomigahara Y, et al. Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. Mol Cell Biol. 2002;22:480–91. doi:10.1128/MCB.22.2.480.

Licht JD. DNA Methylation Inhibitors in cancer therapy: the immunity dimension. Cell. 2015;162:938–9. doi:10.1016/j.cell.2015.08.005.

Lin JC, Jeong S, Liang G, et al. Role of nucleosomal occupancy in the epigenetic silencing of the MLH1 CpG island. Cancer Cell. 2007;12:432–44. doi:10.1016/j.ccr.2007.10.014.

Lou S, Lee H-M, Qin H, et al. Whole-genome bisulfite sequencing of multiple individuals reveals complementary roles of promoter and gene body methylation in transcriptional regulation. Genome Biol. 2014;15:408. doi:10.1186/s13059-014-0408-0.

Lu J, Song G, Tang Q, et al. IRX1 hypomethylation promotes osteosarcoma metastasis via induction of CXCL14/NF- k B signaling. J Clin Invest. 2015;125:1–18. doi:10.1172/JCI78437.Iroquois.

Mair B, Kubicek S, Nijman SMB. Exploiting epigenetic vulnerabilities for cancer therapeutics. Trends Pharmacol Sci. 2014;35:136–45. doi:10.1016/j.tips.2014.01.001.

Meissner A, Mikkelsen TS, Gu H, et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature. 2008;454:766–70. doi:10.1038/nature07107.

Mikkelsen TS, Ku M, Jaffe DB, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 2007;448:553–60.

Moen EL, Stark AL, Zhang W, et al. The role of gene body cytosine modifications in MGMT expression and sensitivity to temozolomide. Mol Cancer Ther. 2014;13:1334–44. doi:10.1158/1535-7163.MCT-13-0924.

Nagarajan RP, Zhang B, Bell RJA, et al. Recurrent epimutations activate gene body promoters in primary glioblastoma. Genome Res. 2014;24:761–74. doi:10.1101/gr.164707.113.

Noushmehr H, Weisenberger DJ, Diefes K, et al. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. Cancer Cell. 2010;17:510–22. doi:10.1016/j.ccr.2010.03.017.

Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell. 1999;99:247–57. doi:10.1016/S0092-8674(00)81656-6.

Oki Y, Aoki E, Issa J-PJ. Decitabine – bedside to bench. Crit Rev Oncol Hematol. 2007;61:140–52. doi:10.1016/j.critrevonc.2006.07.010.

Pastor WA, Aravind L, Rao A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. Nat Rev Mol Cell Biol. 2013;14:341–56. doi:10.1038/nrm3589.

Pastor WA, Pape UJ, Huang Y, et al. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature. 2011;473:394–7.

Pessia E, Makino T, Bailly-Bechet M, et al. Mammalian X chromosome inactivation evolved as a dosage-compensation mechanism for dosage-sensitive genes on the X chromosome. Proc Natl Acad Sci U S A. 2012;109:5346–51. doi:10.1073/pnas.1116763109.

Pfeifer GP. p53 mutational spectra and the role of methylated CpG sequences. Mutat Res. 2000;450:155–66. doi:10.1016/S0027-5107(00)00022-1.

Pontier DB, Gribnau J. Xist regulation and function explored. Hum Genet. 2011;130:223–36. doi:10.1007/s00439-011-1008-7.

Ravenel JD, Broman KW, Perlman EJ, et al. Loss of imprinting of insulin-like growth factor-II (IGF2) gene in distinguishing specific biologic subtypes of wilms tumor. JNCI J Natl Cancer Inst. 2001;93:1698–703. doi:10.1093/jnci/93.22.1698.

Rhee I, Bachman KE, Park BH, et al. DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. Nature. 2002;416:552–6.

Rhee I, Jair KW, Yen RW, et al. CpG methylation is maintained in human cancer cells lacking DNMT1. Nature. 2000;404:1003–7. doi:10.1038/35010000.

Roulois D, Loo Yau H, Singhania R, et al. DNA-demethylating agents target colorectal cancer cells by inducing viral mimicry by endogenous transcripts. Cell. 2015;162:961–73. doi:10.1016/j.cell.2015.07.056.

Saied MH, Marzec J, Khalid S, et al. Genome wide analysis of acute myeloid leukemia reveal leukemia specific methylome and subtype specific hypomethylation of repeats. PLoS One. 2012;7:e33213. doi:10.1371/journal.pone.0033213.

Saito Y, Liang G, Egger G, et al. Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells. Cancer Cell. 2006;9:435–43. doi:10.1016/j.ccr.2006.04.020.

Sakai T, Toguchida J, Ohtani N, et al. Allele-specific hypermethylation of the retinoblastoma tumor-suppressor gene. Am J Hum Genet. 1991;48:880–8.

Sandoval J, Esteller M. Cancer epigenomics: beyond genomics. Curr Opin Genet Dev. 2012;22:50–5. doi:10.1016/j.gde.2012.02.008.

Schmelz K, Wagner M, Dörken B, Tamm I. 5-Aza-2'-deoxycytidine induces p21WAF expression by demethylation of p73 leading to p53-independent apoptosis in myeloid leukemia. Int J Cancer. 2005;114:683–95. doi:10.1002/ijc.20797.

Sharma S, De Carvalho DD, Jeong S, et al. Nucleosomes containing methylated DNA stabilize DNA methyltransferases 3A/3B and ensure faithful epigenetic inheritance. PLoS Genet. 2011;7, e1001286. doi:10.1371/journal.pgen.1001286.

Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. Carcinogenesis. 2010;31:27–36. doi:10.1093/carcin/bgp220.

Shen H, Laird PW. Interplay between the cancer genome and Epigenome. Cell. 2013;153:38–55. doi:10.1016/j.cell.2013.03.008.

Shih AH, Abdel-Wahab O, Patel JP, Levine RL. The role of mutations in epigenetic regulators in myeloid malignancies. Nat Rev Cancer. 2012;12:599–612. doi:10.1038/nrc3343.

Shiohama Y, Ohtake J, Ohkuri T, et al. Identification of a meiosis-specific protein, MEIOB, as a novel cancer/testis antigen and its augmented expression in demethylated cancer cells. Immunol Lett. 2014;158:175–82. doi:10.1016/j.imlet.2014.01.004.

Silber JR, Bobola MS, Blank A, Chamberlain MC. O(6)-methylguanine-DNA methyltransferase in glioma therapy: promise and problems. Biochim Biophys Acta. 2012;1826:71–82. doi:10.1016/j.bbcan.2011.12.004.

Smith ZD, Meissner A. DNA methylation: roles in mammalian development. Nat Rev Genet. 2013;14:204–20. doi:10.1038/nrg3354.

Søes S, Daugaard IL, Sørensen BS, et al. Hypomethylation and increased expression of the putative oncogene ELMO3 are associated with lung cancer development and metastases formation. Oncoscience. 2014;1:367–74.

Srivastava P, Paluch BE, Matsuzaki J, et al. Immunomodulatory action of the DNA methyltransferase inhibitor SGI-110 in epithelial ovarian cancer cells and xenografts. Epigenetics. 2015;10:237–46. doi:10.1080/15592294.2015.1017198.

Stresemann C, Lyko F. Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine. Int J Cancer. 2008;123:8–13. doi:10.1002/ijc.23607.

Su S-F, de Castro Abreu AL, Chihara Y, et al. A panel of three markers hyper- and hypomethylated in urine sediments accurately predicts bladder cancer recurrence. Clin Cancer Res. 2014;20:1978–89. doi:10.1158/1078-0432.CCR-13-2637.

Taniguchi T, Sullivan MJ, Ogawa O, Reeve AE. Epigenetic changes encompassing the IGF2/H19 locus associated with relaxation of IGF2 imprinting and silencing of H19 in Wilms tumor. Proc Natl Acad Sci U S A. 1995;92:2159–63.

TCGA. Integrated genomic analyses of ovarian carcinoma. Nature. 2011;474:609–15. doi:10.1038/nature10166.

Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Poznań, Poland). 2015;19:A68–77. doi:10.5114/wo.2014.47136.

Toyota M. Methylation profiling in acute myeloid leukemia. Blood. 2001;97:2823–9. doi:10.1182/blood.V97.9.2823.

Toyota M, Ahuja N, Ohe-Toyota M, et al. CpG island methylator phenotype in colorectal cancer. Proc Natl Acad Sci U S A. 1999;96:8681–6.

Treppendahl MB, Qiu X, Søgaard A, et al. Allelic methylation levels of the noncoding VTRNA2-1 located on chromosome 5q31.1 predict outcome in AML. Blood. 2012;119:206–16. doi:10.1182/blood-2011-06-362541.

Tsai H-C, Li H, Van Neste L, et al. Transient low doses of DNA-demethylating agents exert durable antitumor effects on hematological and epithelial tumor cells. Cancer Cell. 2012;21:430–46. doi:10.1016/j.ccr.2011.12.029.

Turcan S, Rohle D, Goenka A, et al. IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. Nature. 2012;483:479–83. doi:10.1038/nature10866.

Vrba L, Garbe JC, Stampfer MR, Futscher BW. A lincRNA connected to cell mortality and epigenetically-silenced in most common human cancers. Epigenetics. 2015;10:1074–83. doi:10.1080/15592294.2015.1106673.

Wang J, Chen H, Fu S, et al. The involvement of CHD5 hypermethylation in laryngeal squamous cell carcinoma. Oral Oncol. 2011;47:601–8. doi:10.1016/j.oraloncology.2011.05.003.

Weinstein JN, Collisson EA, Mills GB, et al. The cancer genome atlas pan-cancer analysis project. Nat Genet. 2013;45:1113–20. doi:10.1038/ng.2764.

Weisenberger DJ. Characterizing DNA methylation alterations from the cancer genome atlas. J Clin Invest. 2014;124:17–23. doi:10.1172/JCI69740.

Weisenberger DJ, Liang G. Contributions of DNA methylation aberrancies in shaping the cancer epigenome. Transl Cancer Res. 2015;4:219–34. doi:10.3978/j.issn.2218-676X.2015.05.01.

Weisenberger DJ, Siegmund KD, Campan M, et al. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. Nat Genet. 2006;38:787–93. doi:10.1038/ng1834.

Wolff EM, Byun H-M, Han HF, et al. Hypomethylation of a LINE-1 promoter activates an alternate transcript of the MET oncogene in bladders with cancer. PLoS Genet. 2010;6:e1000917. doi:10.1371/journal.pgen.1000917.

Wrangle J, Wang W, Koch A, et al. Alterations of immune response of non-small cell lung cancer with azacytidine. Oncotarget. 2013;4:2067–79.

Wu MS, Wang HP, Lin CC, et al. Loss of imprinting and overexpression of IGF2 gene in gastric adenocarcinoma. Cancer Lett. 1997;120:9–14.

Yamazaki J, Issa J-PJ. Epigenetic aspects of MDS and its molecular targeted therapy. Int J Hematol. 2013;97:175–82. doi:10.1007/s12185-012-1197-4.

Yang X, Han H, De Carvalho DD, et al. Gene body methylation can alter gene expression and is a therapeutic target in cancer. Cancer Cell. 2014;26:577–90. doi:10.1016/j.ccr.2014.07.028.

Yang X, Lay F, Han H, Jones PA. Targeting DNA methylation for epigenetic therapy. Trends Pharmacol Sci. 2010;31:536–46. doi:10.1016/j.tips.2010.08.001.

Yegnasubramanian S, Haffner MC, Zhang Y, et al. DNA hypomethylation arises later in prostate cancer progression than CpG island hypermethylation and contributes to metastatic tumor heterogeneity. Cancer Res. 2008;68:8954–67. doi:10.1158/0008-5472.CAN-07-6088.

Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet. 1997;13:335–40.

Yoo CB, Jeong S, Egger G, et al. Delivery of 5-aza-2'-deoxycytidine to cells using oligodeoxynucleotides. Cancer Res. 2007;67:6400–8. doi:10.1158/0008-5472.CAN-07-0251.

You JS, Jones PA. Cancer genetics and epigenetics: two sides of the same coin? Cancer Cell. 2012;22:9–20. doi:10.1016/j.ccr.2012.06.008.

Zarnett OJ, Sahgal A, Gosio J, et al. Treatment of elderly patients with glioblastoma: a systematic evidence-based analysis. JAMA Neurol. 2015;72:589–96. doi:10.1001/jamaneurol.2014.3739.

Structure and Mechanism of Plant DNA Methyltransferases

Jiamu Du

Abstract
DNA methylation is an important epigenetic mark that functions in eukaryotes from fungi to animals and plants, where it plays a crucial role in the regulation of epigenetic silencing. Once the methylation mark is established by the de novo DNA methyltransferase (MTase), it requires specific regulatory mechanisms to maintain the methylation state during chromatin replication, both during meiosis and mitosis. Plants have distinct DNA methylation patterns that are both established and maintained by unique DNA MTases and are regulated by plant-specific pathways. This chapter focuses on the exceptional structural and functional features of plant DNA MTases that provide insights into these regulatory mechanisms.

Abbreviations

5mC	5-Methyl-cytosine
6mA	6-Methyl-adenine
AdoHcy	S-Adenosyl-L-homocysteine
AdoMet	S-Adenosyl-L-methionine
AGO4	ARGONAUTE 4
BAH domain	Bromo-adjacent homology domain
CMT	CHROMOMETHYLASE
DCL3	DICER-LIKE 3

J. Du
Shanghai Center for Plant Stress Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 201602, China
e-mail: jmdu@sibs.ac.cn

© Springer International Publishing Switzerland 2016

A. Jeltsch, R.Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*, Advances in Experimental Medicine and Biology 945,
DOI 10.1007/978-3-319-43624-1_8

DDM1	DECREASED IN DNA METHYLATION 1
DRM2	DOMAINS REARRANGED METHYLTRANSFERASE 2
ITC	Isothermal titration calorimetry
KYP	KRYPTONITE
MET1	DNA METHYLTRANSFERASE 1
MTase	Methyltransferase
Pol II/IV/V	RNA polymerase II/IV/V
RdDM	RNA-directed DNA methylation
RDR2	RNA-DEPENDENT RNA POLYMERASE 2
RFTD	Replication foci targeting domain
SET domain	Su(var)3-9, enhancer of zeste, trithorax domain
SHH1	SAWADEE HOMEODOMAIN HOMOLOG 1
SRA domain	SET and RING finger-associated domain
SUVH	SUPPRESSOR OF VARIEGATION 3-9 HOMOLOG
TE	Transposable elements
TRD	Target recognition domain
UBA	Ubiquitin-associated domain
UHRF1	Ubiquitin-like PHD and RING finger domains 1
VIM	VARIANT IN METHYLATION
ZMET2	Zea methyltransferase 2

1 Introduction

The DNA methylation modification represents the addition of a methyl group to DNA. In plants, DNA methylation can occur both at the 5 position of the cytosine base (5mC) and at the 6 position of the adenine base (6mA) (Vanyushin and Ashapkin 2011). 5mC has been extensively studied as an important part of bacterial defense systems and eukaryotic epigenetic regulation systems ranging from fungi to humans, even though several species, such as yeast and *Caenorhabditis elegans*, lack this type of epigenetic mark (Goll and Bestor 2005). In contrast to that, the 6mA has been less studied. Thus, in a narrow sense, DNA methylation usually only stands for the 5mC mark, and this chapter will only focus on 5mC-related studies in higher plants. Studies from fungi to plants and higher mammalian systems have established the conserved function of DNA methylation in gene silencing, genome imprinting, and the repression of transposable elements (TE) and repeat sequences (Law and Jacobsen 2010; Castel and Martienssen 2013). DNA methylation requires specific enzymes, namely, DNA methyltransferases (MTases), which share a common catalytic mechanism that enables transfer of a methyl group from the methyl donor S-adenosyl-methionine (AdoMet) to the 5 position of the cytosine base. Nevertheless, these enzymes differ in their sequence specificity and regulatory mechanisms. Plants have evolved in a distinct manner from animals and hence plants invoke some specific regulatory pathways controlling DNA methylation. In this chapter, the focus will be on the structure and mechanism of plant DNA MTases, thereby highlighting the unique DNA methylation system of plants. As most of the

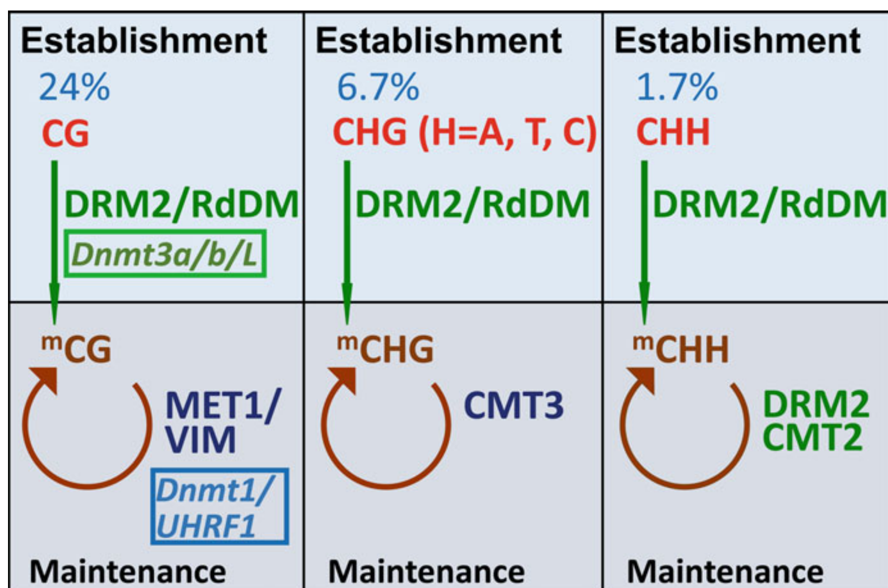


Fig. 1 Establishment and maintenance of plant DNA methylation and the corresponding MTases. Plants possess three types of DNA methylation patterns: CG, CHG, and CHH, all of which are established by DRM2 (an ortholog of mammalian Dnmt3a) driven by the RdDM pathway. The maintenance of CG methylation requires MET1 and its cofactor VIM proteins, which are orthologs of mammalian Dnmt1 and UHRF1, respectively. CHG methylation is mainly maintained by plant-specific DNA MTase CMT3. CHH methylation is maintained through two pathways: the DRM2 pathway driven by RdDM and CMT2-mediated CHH methylation

plant DNA methylation studies are carried out using the model system *Arabidopsis thaliana*, the emphasis will be on the use of *Arabidopsis* genes to present current studies on plant DNA methylation, unless specified otherwise.

Distinct from symmetrical DNA methylation at CG dinucleotide sites which dominates in mammals, plant DNA methylation is much more complicated and occurs both symmetrically and asymmetrically in all three sequence contexts, CG, CHG (H denotes A, T, or C), and CHH, with methylation levels of about 24 %, 6.7 %, and 1.7 % in *Arabidopsis*, respectively (Cokus et al. 2008) (Fig. 1). In pericentromeric heterochromatin and some small patches in euchromatin region, all the three types of DNA methylation are heavily distributed over the TE and repeat sequences (Lister et al. 2008; Cokus et al. 2008). The heterochromatic DNA methylation has extensive internal cross talk and is shown to highly correlate with silent histone marks such as histone 3 lysine 9 dimethylation (H3K9me2) and H3K4me0 (Du et al. 2015). It serves to silence such elements to preserve genome integrity and to act as a genomic immune system (Law and Jacobsen 2010; Kim and Zilberman 2014; Matzke and Mosher 2014). In contrast, CG methylation is observed not only abundantly in heterochromatic regions for repression of TE and repeat sequences but also with a less extent in euchromatic genic regions in one third of the transcribed genes for regulation of gene silencing and activation (Castel and Martienssen 2013; Law and Jacobsen 2010; Cokus et al. 2008; Lister et al. 2008). In plant, CG

methylation is independent of the silencing mark H3K9me2 and shows limited cross talk with non-CG methylation (Du et al. 2015).

All de novo methylation in plants utilizes a plant-specific RNA-directed DNA methylation (RdDM) pathway to guide the de novo DNA MTase DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) (Fig. 1), which is an ortholog of mammalian Dnmt3a, but with a significant rearrangement of MTase signature motifs (Cao and Jacobsen 2002a, b). The maintenance of different DNA methylation patterns in plants requires the employment of different pathways. CG methylation is maintained through a similar pathway as in mammals. The DNA METHYLTRANSFERASE 1 (MET1) which is an ortholog of the mammalian Dnmt1 maintains plant CG methylation in cooperation with the VARIANT IN METHYLATION (VIM) proteins, which are orthologs of mammalian ubiquitin-like PHD and RING finger domains 1 (UHRF1) (Fig. 1) (Finnegan and Kovac 2000; Woo et al. 2008). Most CHG methylation is maintained by a self-reinforcing loop between the plant-specific CHG DNA MTase CHROMOMETHYLASE 3 (CMT3) and the H3K9 MTase KRYPTONITE [KYP, also known as SUPPRESSOR OF VARIEGATION 3–9 HOMOLOG 4 (SUVH4)] and its close homologs SUVH5 and SUVH6 (Du et al. 2012, 2014, 2015; Jackson et al. 2002; Lindroth et al. 2001). CHH methylation is totally asymmetric so that the maintenance of CHH methylation biochemically equals de novo methylation, which drives DRM2 to maintain CHH methylation through RdDM (Law and Jacobsen 2010). An alternative pathway, CMT2-controlled CHH methylation, was shown to be responsible for the majority of heterochromatic CHH methylation in *Arabidopsis* (Fig. 1), which can form a similar self-reinforcing loop with KYP as CMT3 (Stroud et al. 2014; Zemach et al. 2013). However, the plant DNA methylation is not a simple system in which each enzyme performs its own exclusive function. There is extensive cross talk between different methylation patterns, especially between the two types of non-CG methylation. CMT2 functions in the maintenance of CHG methylation at some loci supporting CMT3, while CMT3 also functions in CHH methylation together with CMT2 and DRM2 in a redundant manner (Stroud et al. 2014). Altogether, there are four types of functionally active DNA MTases in *Arabidopsis*: DRM2, MET1, CMT3, and CMT2. The structural features and molecular mechanisms of each of these plant DNA MTases will be discussed in the following sections. For details about mammalian and bacterial enzymes refer to other chapters in this book.

2 Structure and Mechanism of Plant DNA MTases

2.1 Structural Mechanism of the Maintenance of CHG Methylation in Plant

2.1.1 Overview of Plant CHG DNA Methylation

CMT type DNA MTases are evolutionary conserved plant-specific DNA MTases that have not been identified in species other than plants. They feature an N-terminal bromo-adjacent homology (BAH) domain and a conserved chromodomain

embedded within the C-terminal DNA MTase domain (Fig. 2a). *Arabidopsis* CMT3 was first cloned by a forward genetic screen of suppressor genes in the hypermethylated *clark kent* mutant strain background, with the *cmt3* mutants showing a significant decrease of genome-wide CHG methylation levels (Lindroth et al. 2001). The chromodomain has been studied extensively as a histone methyllysine reader module (Blus et al. 2011), providing a potential linkage between readout of methylated histone and the establishment of DNA methylation. Shortly after the discovery of CMT3, another gene was identified to be critical for CHG methylation using the same screening system, which turned out to be an SUVH family histone MTase and was subsequently named KRYPTONITE (Jackson et al. 2002). The SUVH family histone MTases share a common architecture consisting of an N-terminal SET and RING finger-associated (SRA) domain, which is capable of recognizing methylated DNA, and a C-terminal Su(var)3-9, enhancer of zeste, trithorax (SET) domain which is a putative histone MTase domain (Fig. 2b), showing that methylated DNA may play a role in histone methylation (Johnson et al. 2007, 2008). Altogether, a self-reinforcing loop model was postulated to link H3K9me2 and DNA methylation by cycling between CMT3 and KYP (Law and Jacobsen 2010) (Fig. 2c).

2.1.2 Structure and Mechanism of CMT3

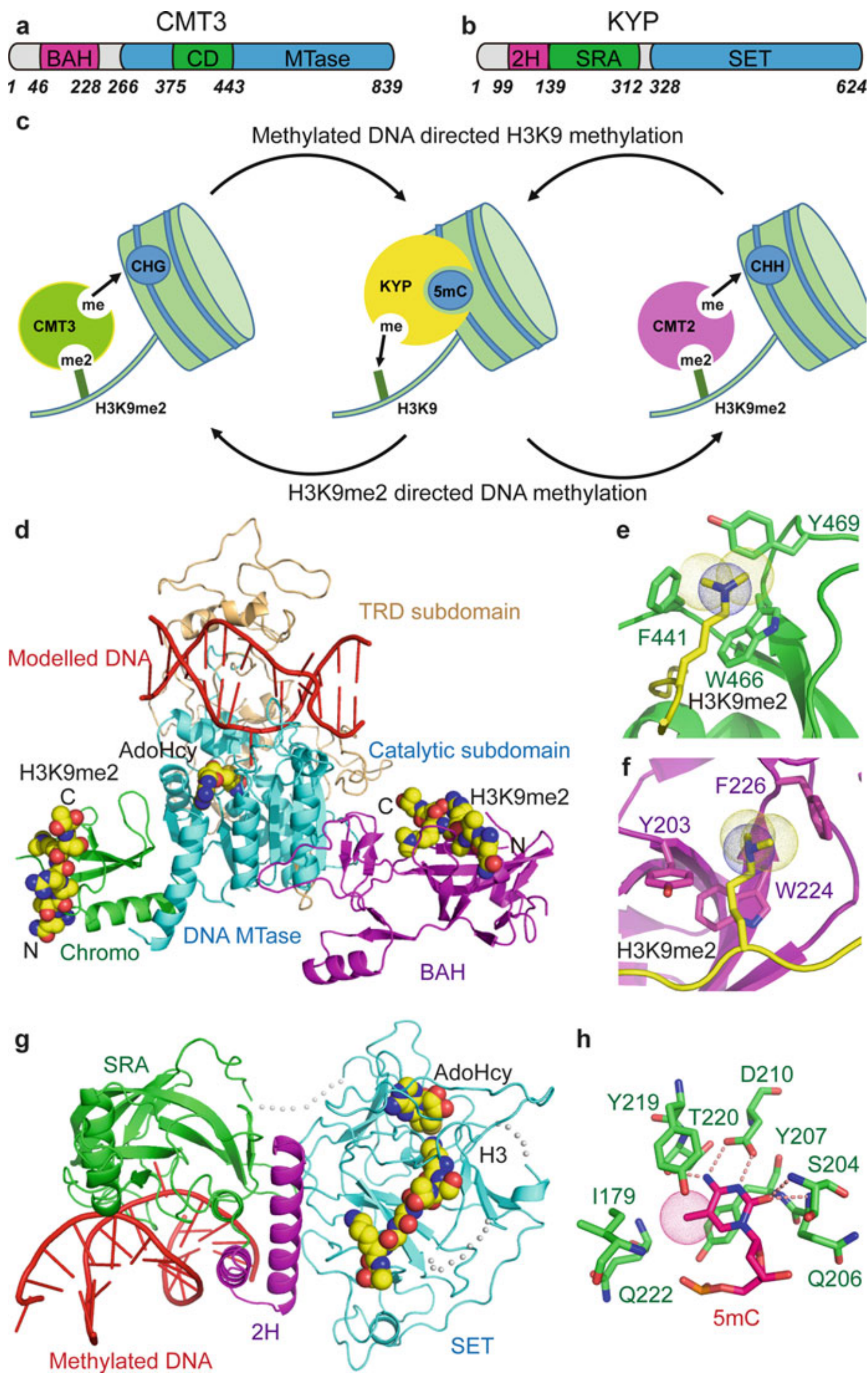
The co-occurrence between CHG DNA methylation and the H3K9me2 histone mark has been observed throughout the genome in plants (Bernatavichute et al. 2008), raising the possibility of the genome-wide association between CMT3 and H3K9me2. This was subsequently confirmed by both in vitro pulldown and whole-genome chromatin immunoprecipitation coupled with sequencing (ChIP-seq) experiments (Du et al. 2012). The direct binding of CMT3 to H3K9me2 was postulated because of the existence of the chromodomain and was confirmed by in vitro peptide chip array containing hundreds of different combinations of histone modifications (Du et al. 2012). Interestingly, quantitative measurements of the binding between CMT3 or its maize equivalent Zea methyltransferase 2 (ZMET2), with an H3(1-15)K9me2 peptide by isothermal titration calorimetry (ITC), yielded a stoichiometry of around 2, indicating that two H3K9me2 binding sites exist on each CMT3 or ZMET2 molecules (Du et al. 2012; Stroud et al. 2014). The ITC with individual domains of ZMET2 further revealed that the BAH and chromo domains correspond to the two binding sites for H3K9me2 on CMT3/ZMET2 (Du et al. 2012).

Although crystallization attempts with CMT3 failed, the crystal structure of ZMET2 in complex with bound cofactor S-adenosyl-L-homocysteine (AdoHcy) was successfully determined and provided the first structure of a plant DNA MTase (Fig. 2d) (Du et al. 2012). The structure of the AdoHcy-bound form of ZMET2 containing all the functional domains (BAH, chromo, and DNA MTase) adopts a triangular-shaped fold (Fig. 2d). The BAH and chromo domains are positioned on the two triangular-like edges of the MTase domain, despite the chromodomain being embedded inside the MTase domain in the primary sequence (Fig. 2a, d) (Du et al. 2012). The BAH and chromo domains, together with the target recognition subdomain (TRD) of the MTase domain, form the three vertices of a triangular-shaped topology, with the catalytic pocket located in the center of the triangle

(Du et al. 2012), revealing a potential regulatory mechanism of MTase activity by the surrounding domains. It is worth noting that the relative positioning of the ZMET2 BAH domain against the MTase domain resembles the relative positioning observed for the first BAH domain and the MTase domain of mouse Dnmt1, suggesting a common evolutionary origin and plausibly a similar regulatory role of the BAH domain (Du et al. 2012; Song et al. 2011). The MTase domain adopts a classical type I MTase fold composed of a catalytic subdomain and a TRD subdomain (Fig. 2d). The catalytic subdomain adopts the classic sandwich topology with a central seven-stranded β-sheet flanked by two layers of α-helices on either side (Fig. 2d), which resembles other reported structures of type I DNA MTases, such as M.HhaI, Dnmt3a, and Dnmt1 (Jia et al. 2007; Song et al. 2011; Du et al. 2012; Cheng et al. 1993). The TRD subdomain adopts a novel folding topology enriched with loops, which has never been observed in structures of other DNA MTases (Fig. 2d), suggesting a plausible novel DNA recognition model, although the structure with bound DNA is lacking at this time (Du et al. 2012). The BAH and chromo domains are firmly anchored on the two edges of the MTase domain by extensive interdomain interactions (Du et al. 2012). Two β-strands of the BAH domain form a continuous nine-stranded β-sheet with the central seven-stranded β-sheet of the MTase domain which both stabilizes and defines the relative position of the BAH domain (Du et al. 2012). The chromodomain has several hydrophobic residues to form a hydrophobic core with the MTase domain, thereby stabilizing their relative positions (Du et al. 2012). Therefore, the BAH and chromo domains adopt a rigid alignment against the MTase domain and possess a potential regulatory role to modulate the MTase domain function.

Further structural studies of ZMET2-H3K9me2 complexes revealed a regulatory mechanism that directs CHG methylation in plants by the silencing mark H3K9me2 (Du et al. 2012). The crystal structure of ZMET2-AdoHcy-H3(1-15)

Fig. 2 Structure and mechanism of maintenance of CHG methylation in plant. (**a**) Schematic representation of the domain architecture of CMT3. CD, chromodomain. (**b**) Schematic representation of the domain architecture of KYP. 2H, two-helix bundle. (**c**) A schematic model showing that methylated DNA-directed H3K9 methylation by KYP can form self-reinforcing loops with H3K9me2-directed CHG DNA methylation by CMT3 or CHH DNA methylation by CMT2. (**d**) A structural model of ZMET2 in complex with the cofactor AdoHcy, two H3K9me2 peptides, and a modeled DNA (PDB codes: 4FSX, 4FT2, and 4FT4. The DNA was modeled based on PDB: 4DA4). The BAH domain, catalytic subdomain, TRD subdomain, chromodomain, and the modeled DNA are colored in *magenta*, *cyan*, *orange*, *green*, and *red*, respectively. The two peptides are shown in space-filling representation with their C-termini directed toward the catalytic center. (**e**) Three aromatic residues (F441, W466, and Y469) of ZMET2 chromodomain form an aromatic cage to specifically recognize the methyllysine of H3K9me2. (**f**) Three aromatic residues (Y203, W224, and F226) of ZMET2 BAH domain form an aromatic cage to specifically recognize the methyllysine of H3K9me2. (**g**) Crystal structure of KYP in complex with methylated DNA, AdoHcy, and H3 peptide. The two-helix bundle (2H), SRA domain, SET domain, and the methylated DNA are colored in *magenta*, *green*, *cyan*, and *red*, respectively. The AdoHcy and H3 peptide are shown in space-filling representation. (**h**) Structural basis for the specific recognition of 5mC by a small pocket within the SRA domain


a
CMT3
BAH
CD
MTase
1 46 228 266 375 443 839
b
KYP
2H
SRA
SET
1 99 139 312 328 624
c
Methylated DNA directed H3K9 methylation
CHG
me
CMT3
me2
H3K9me2
KYP
5mC
me
H3K9
CHH
me
CMT2
me2
H3K9me2
H3K9me2 directed DNA methylation
d
TRD subdomain
Modelled DNA
AdoHcy
Catalytic subdomain
H3K9me2
C
C
H3K9me2
N
N
Chromo
DNA MTase
BAH
e
Y469
F441
W466
H3K9me2
f
F226
Y203
W224
H3K9me2
g
SRA
AdoHcy
H3
2H
SET
Methylated DNA
h
D210
Y219
T220
Y207
S204
I179
Q206
Q222
5mC

K9me2 complex adopts a conformation whereby the H3K9me2 peptide binds to the chromodomain (Du et al. 2012). Upon binding by the peptide, the ZMET2 protein shows a nearly identical conformation as in its free form, revealing rigid binding without allosteric regulation. The chromodomain recognizes the peptide using a classic recognition mode in which the peptide adopts an extended β-strand-like conformation stabilized through intermolecular main chain hydrogen-bonding interactions with the chromodomain (Du et al. 2012; Blus et al. 2011). The dimethyllysine side chain inserts into a classic aromatic cage formed by three aromatic residues like that observed for other chromodomains and is stabilized through both hydrophobic and cation-π interactions (Fig. 2e) (Du et al. 2012; Blus et al. 2011; Patel and Wang 2013). The peptide possesses a specific directionality with its C-terminus directed toward the catalytic center of the MTase domain (Fig. 2d), revealing a plausible mechanism that positions the catalytic center of MTase domain toward the inner core region of the nucleosome.

In the ZMET2-AdoHcy-H3(1-32)K9me2 complex, the peptide is bound to the BAH domain (Du et al. 2012). The BAH domain has been extensively studied both structurally and functionally as a histone methyllysine reader module (Yang and Xu 2013; Kuo et al. 2012), though this is the first report of BAH domain recognition of an H3K9me2 mark (Yang and Xu 2013; Du et al. 2012). The ZMET2 BAH domain forms less main chain interactions with the H3K9me2 peptide compared with the chromodomain. However, it also contains an aromatic cage formed by three aromatic residues for proper accommodation of the dimethyllysine side chain through hydrophobic and cation-π interactions, similar to other methyllysine reader modules (Fig. 2f) (Du et al. 2012; Patel and Wang 2013). Interestingly, the BAH domain-bound peptide has similar directionality as that observed for the chromodomain-bound peptide, and its C-terminus is directed toward the catalytic center of the MTase domain (Fig. 2d), revealing a similar regulation mechanism as the chromodomain, thereby positioning the MTase domain toward the inner core region of the nucleosome.

The observation that the histone tails bound by both BAH and chromo domains have the same directionality raises the possibility that the two domains coordinate the targeting of the MTase domain in a combinatorial manner. Interestingly, triple mutants of the three aromatic cage residues of either the BAH or chromo domains of CMT3 disrupting the binding capacity to H3K9me2 by these reader modules retain their in vitro MTase activity but fail to complement the in vivo function of CMT3 in *cmt3* mutation background (Du et al. 2012), suggesting that the H3K9me2 recognition capacity of both domains is essential for its in vivo function. Therefore, the BAH and chromo domains need to simultaneously target the MTase domain onto the H3K9me2-containing nucleosome. The two H3K9me2 binding sites within the BAH and chromo domains are separated by a distance of about 70 Å which can span the diameter of a single nucleosome, allowing a model whereby the CMT3 can bind to a single nucleosome, such that the two H3K9me2 tails can extend out from the nucleosome core region and are captured by the BAH and chromo domains of CMT3 (Du et al. 2012). In this model, the CMT3 will preferentially methylate the local DNA nearby the H3K9me2-containing nucleosome.

Alternatively, since the H3K9me2 mark is enriched in heterochromatin regions, it is possible that the BAH and chromo domains of CMT3 protein are targeted on adjacent nucleosomes to allow the CMT3 to "walk" throughout the heterochromatin region using BAH and chromo domains as two "legs," representing a DNA methylation spreading mechanism. CHG methylation is mainly distributed in heterochromatin regions, which are highly overlapping with the H3K9me2-enriched regions, thereby causing the silencing of TE in plants. The strict requirement of the two H3K9me2 marks for in vivo function of CMT3 can ensure that the CMT3 is targeted to proper H3K9me2-enriched regions such as heterochromatic regions but not randomly silence other CHG sites.

2.1.3 Structure and Mechanism of KRYPTONITE

Since CHG methylation solely depends on H3K9me2, the maintenance of H3K9me2 is necessary for proper maintenance of CHG methylation during mitosis and meiosis. Thus, it is indispensable to study H3K9 methylation during investigation of CHG methylation in plants. In *Arabidopsis*, H3K9 methylation relies on the SUVH family H3K9 MTase KYP and its homologs SUVH5 and SUVH6. A *kyp suvh5 suvh6* triple mutant strain which has lost most of the H3K9me2 mark in a genome-wide scale shows a significant loss of CHG methylation similar to the *cmt3* mutant strain (Stroud et al. 2013), confirming that CHG methylation is also controlled by H3K9 MTases. The crystal structure of KYP in complex with its cofactor product AdoHcy, methylated CHH or CHG site (mCHH/mCHG) containing DNA, and the substrate H3(1-15) peptide highlights how the H3K9 MTase is regulated by methylated DNA (Fig. 2g) (Du et al. 2014). KYP contains an N-terminal two-helix bundle, a middle SRA domain which can recognize methylated CHH or CHG DNA (Johnson et al. 2007), and a C-terminal SET domain (including pre-SET, SET, and post-SET subdomains) which constitute the H3K9 MTase activity (Jackson et al. 2004). The two-helix bundle is located within the middle of the structure holding together the whole architecture of the protein (Du et al. 2014). The SRA domain is aligned on one side of the two-helix bundle, while the SET domain is aligned on the other side (Fig. 2g) (Du et al. 2014). There are extensive interactions between the three segments of the structure, suggesting a rigid alignment of the entire topology (Du et al. 2014). The KYP SRA domain possesses a positively charged surface cleft and holds the bound DNA, similar to structures of other reported DNA-bound SRA domains (Hashimoto et al. 2008; Arita et al. 2008; Avvakumov et al. 2008; Rajakumara et al. 2011). In addition, some positively charged residues of the two-helix bundle are involved in recognition of the backbone of the methylated DNA (Du et al. 2014). The 5mC base is flipped out from the DNA duplex and inserts into a small pocket of the SRA domain, thereby forming extensive intermolecular interactions with the surrounding residues (Fig. 2h) (Du et al. 2014). Mutations of the 5mC-binding pocket residues lead to a loss of the capacity to bind methylated DNA while retaining the in vitro histone MTase activity. Interestingly, these mutations impair the in vivo function of KYP, suggesting that specific binding to methylated DNA is required for the proper targeting of KYP (Du et al. 2014). The cofactor product AdoHcy and the substrate peptide are clamped in between the SET and post-SET

subdomains (Du et al. 2014). Several important tyrosine residues are positioned around the target lysine and form hydrogen-bonding interactions that can facilitate the catalytic reaction and restrict the enzyme to be an H3K9me2 MTase (Du et al. 2014). Overall, the structures of the KYP-mCHH/CHG DNA-AdoHcy-H3 peptide complex and the DNA-free form of the KYP homolog SUVH9 are nearly identical, indicating neither conformational change nor allosteric regulation upon DNA binding (Du et al. 2014; Johnson et al. 2014). Therefore, the KYP protein is recruited by the methylated CHH or CHG DNA and then methylates nearby histone tails.

In summary, CMT3 uses its BAH and chromo domains to target its DNA MTase domain to H3K9me2-containing nucleosome to achieve H3K9me2-directed CHG DNA methylation. On the contrary, KYP uses its SRA domain to target its histone MTase domain to mCHG/mCHH-containing nucleosomes to process the methylated DNA-directed H3K9 methylation. Taken together, the two proteins form a simple self-reinforcing feedback loop to strengthen H3K9 and CHG methylation at heterochromatin to keep the silenced state of TE and repeat sequences.

2.2 Mechanism of CMT2-Mediated CHH Methylation

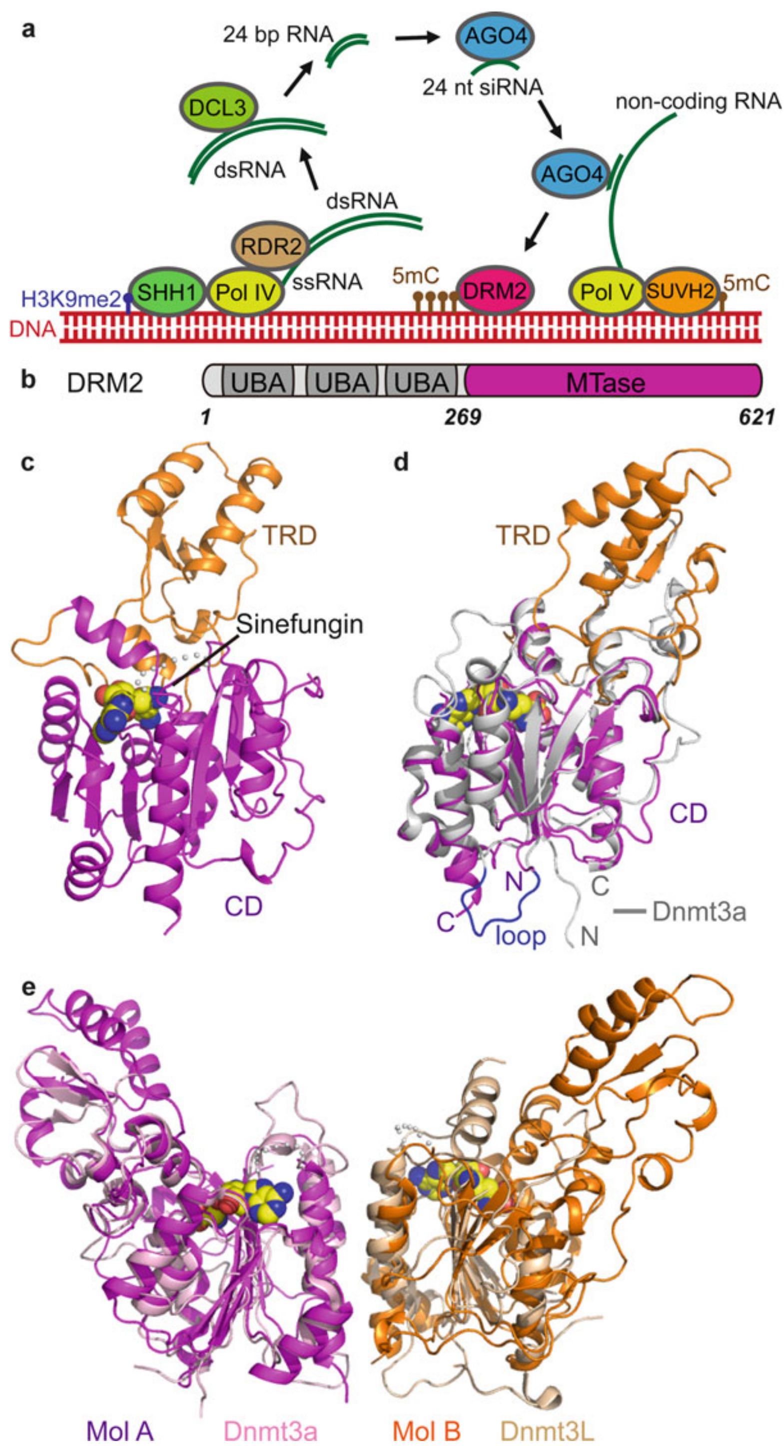
CMT2 was firstly identified as a downstream effector MTase of the chromatin remodeler DECREASED IN DNA METHYLATION 1 (DDM1)-mediated CHH methylation of TE, which is independent of the classical RdDM-dependent CHH methylation (Zemach et al. 2013). A further biochemical characterization of CMT2 revealed that this MTase can de novo methylate both CHH and CHG sites in vitro, in contrast to CMT3 that prefers to maintain CHG site methylation (Stroud et al. 2014). A *cmt2 cmt3* double mutant strain showed a stronger loss of CHG methylation than the *cmt3* mutant alone, confirming that the in vitro CHG methylation activity of CMT2 is also functionally relevant in vivo (Stroud et al. 2014). In addition, the *drm1 drm2 cmt2* triple mutant can nearly completely eliminate CHH methylation genome wide, while the *drm1 drm2* double mutant or *cmt2* single mutant can only partially reduce CHH methylation, indicating that the RdDM pathway and CMT2 are complementary in controlling almost all the CHH methylation (Stroud et al. 2014). Similar to CMT3, the *kyp suvh5 suvh6* triple mutant, which can eliminate H3K9me2, shows 86 % loss of CHH methylation controlled by CMT2, revealing a linkage between H3K9me2 and CMT2 (Stroud et al. 2014). CMT2 contains a middle BAH domain and C-terminal chromodomain inserted within the MTase domain, with a similar domain architecture as CMT3. Thus, it is reasonable to speculate that CMT2 acts in a similar way as CMT3. ITC binding experiments established the CMT2 can directly bind H3K9me with a preference for H3K9me2 and reduced binding of H3K9me1 and H3K9me3, which is consistent with the observation that CMT2 controlled targets have a higher H3K9me2 level than H3K9me1 (Stroud et al. 2014). The ITC binding yields a protein to peptide molar ratio of 1:2, establishing that the dual recognition mode observed for CMT3 is also conserved for CMT2 (Stroud et al. 2014). Due to a high-sequence homology between CMT3 and CMT2 (46 % identity over the functional BAH, chromo, and MTase domains),

it is speculated that CMT2 can use its BAH and chromo domains to target the MTase domain to H3K9me2-containing nucleosomes to achieve position-specific CHH methylation, forming a self-reinforcing loop with KYP, similar to that observed for CMT3 (Fig. 2c). CMT2 has an extra-long N-terminal extension of about 500 residues without any homology to known domains and without predictable secondary structure, and it is not apparent if this segment has any functional role within CMT2. Unlike CMT3, which exists in almost all plant species, CMT2 exists in some but not all plant species. For example, crop maize does not have the homologous gene of CMT2, while crop rice possesses a single CMT2 homologous gene (Zemach et al. 2013). Therefore, CMT2 controlled CHH methylation may not be a strictly conserved plant DNA methylation pathway, in contrast to DRM2 driven by RdDM. Moreover the CMT2 pathway is functionally partially redundant with RdDM. Thus, CMT2 can be kept in an RdDM redundant way in some plant species, while in others it could have been superseded by RdDM during evolution. Functionally, CMT2 was identified to associate with the heat response by comparison with the genome of different collections of *Arabidopsis* (Shen et al. 2014). The *cmt2* mutant has a stronger tolerance to heat-stress, indicating an epigenetic basis for the adaptation to environmental stress (Shen et al. 2014).

2.3 RNA-Directed DNA Methylation (RdDM)

2.3.1 Overview of RdDM

RdDM is a plant-specific de novo DNA methylation pathway, which conducts all the de novo DNA methylation in plants and is required for the maintenance of CHH methylation within small euchromatic sites (He et al. 2014; Matzke and Mosher 2014; Zhao and Chen 2014). The pathway starts with a plant-specific RNA polymerase IV (Pol IV), which is a specialized RNA polymerase that evolved from Pol II. Pol IV is targeted to certain loci to generate single-stranded RNA transcripts (Fig. 3a) (Haag and Pikaard 2011). The RNA transcripts are subsequently used as templates by Pol IV-associated RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) to generate double-stranded RNA (Fig. 3a) (Haag et al. 2012). The double-stranded RNA is then cut into 24 bp siRNAs by DICER-LIKE 3 (DCL3) and one 24-nt strand is further loaded into ARGONAUTE 4 (AGO4) (Fig. 3a) (Law and Jacobsen 2010; Matzke and Mosher 2014). Meanwhile, another plant-specific RNA polymerase Pol V can be targeted to certain loci and produce long noncoding scaffold RNA transcript. The AGO4-bound siRNA can interact with Pol V transcripts by base-pairing as well as protein interaction and further target the plant-specific DNA MTase DRM2 to facilitate site-specific DNA methylation (Fig. 3a) (Law and Jacobsen 2010; Matzke and Mosher 2014). Thus, in this pathway, DNA methylation sites on the chromatin are determined by the targeting of Pol IV, Pol V, and DRM2. Recent studies have established additional mechanisms in which SAWADEE HOMEODOMAIN HOMOLOG 1 (SHH1, also called DNA-binding transcription factor 1, DTF1) can target Pol IV to H3K9me2 containing chromatin regions and the catalytically inactive SUVH family proteins SUVH2/9 can direct



a
24 bp RNA
AGO4
24 nt siRNA
DCL3
non-coding RNA
dsRNA
AGO4
dsRNA
RDR2
5mC
ssRNA
SHH1
Pol IV
DRM2
Pol V
SUVH2
5mC
H3K9me2
DNA
b
DRM2
UBA
UBA
UBA
MTase
1
269
621
c
TRD
Sinefungin
CD
d
TRD
CD
N
C
C
loop
N
Dnmt3a
e
Mol A
Dnmt3a
Mol B
Dnmt3L

Pol V to methylated DNA-containing loci (Fig. 3a) (Law et al. 2013; Johnson et al. 2014; Zhang et al. 2013; Liu et al. 2014). Co-IP experiments indicated that DRM2 occurs in the same complex with AGO4, suggesting that AGO4 plays an important role in the targeting of DRM2 (Zhong et al. 2014). However, the molecular mechanism of the AGO4-DRM2 interaction remains unclear.

2.3.2 Structure and Mechanism of DRM2

Arabidopsis DRM2 was identified as the de novo MTase in plants by a genetic screen and showed sequence conservation with the mammalian de novo MTase Dnmt3a (Cao and Jacobsen 2002a, b). From a sequence perspective, DRM family proteins have N-terminal 2–3 ubiquitin-associated domains (UBAs) and a C-terminal DNA MTase domain which possesses all the signature motifs of type I DNA MTase, except that an alignment established rearrangement of these motifs (Fig. 3b) (Cao and Jacobsen 2002a). This type of unique rearrangement only exists in plants and is considered as a plant-specific feature. *Arabidopsis* DRM2 has no detectable enzymatic activity in vitro (Zhong et al. 2014), revealing that it may need some cofactors for its activity. In contrast, the DRM from *Nicotiana tabacum* (NtDRM) has de novo DNA MTase activity at CHH and CHG sites both in vitro and in vivo (Wada et al. 2003). The structure of NtDRM MTase domain indicates a classic type I DNA MTase fold with a central seven-stranded β-sheet flanked by two layers of α-helices positioned on both sides (Fig. 3c) (Zhong et al. 2014). The TRD subdomain of NtDRM is composed of a two-stranded antiparallel β-sheet and two antiparallel α-helices, which are different from structures of other reported DNA MTases, indicative of a unique DNA recognition and selection mode (Fig. 3c) (Zhong et al. 2014). The catalytic subdomain is similar to other DNA MTase domains such as the Dnmt3a MTase domain (Jia et al. 2007; Zhong et al. 2014). In the Dnmt3a MTase domain crystal structure, the N- and C-termini are close to each other (Jia et al. 2007). If the N- and C-termini of Dnmt3a MTase fused together and a break is incorporated at the loop between Pro739 and Pro746 (highlighted in blue in Fig. 3d) (Jia et al. 2007), the resulting sequence-based folding topology would be identical

Fig. 3 Structure and mechanism of DRM2 function driven by RdDM. (**a**) A schematic model of RdDM pathway. Pol IV is targeted by SHH2 to H3K9me2-containing loci to produce ssRNA transcripts. The ssRNA is doubled by Pol IV-associated RDR2 and cleaved by DCL3 to generate 24-bp dsRNA. The 24-nt siRNA is then loaded onto AGO4. Pol V is guided by SUVH2 (or SUVH9) to methylated DNA-containing loci to generate scaffold noncoding RNA transcripts. The noncoding RNA can interact with the AGO4-bound siRNA to further direct DRM2 to perform de novo DNA methylation. (**b**) Domain architecture of *Arabidopsis* DRM2. (**c**) Crystal structure of the catalytic domain of NtDRM with the catalytic subdomain (CD) and TRD subdomain colored in *magenta* and *orange*, respectively (PDB code: 4ONJ). (**d**) Superposition of the NtDRM catalytic domain with the Dnmt3a catalytic domain, which is colored in *silver* (PDB codes: 4ONJ and 2QRV). If the N and C-termini of Dnmt3a are fused together and a break incorporated within the loop marked as blue, the overall topology of Dnmt3a becomes similar to DRM. (**e**) NtDRM catalytic domain shows a dimeric arrangement with one molecule colored in *magenta* and another in *orange*. The DRM homodimer interface mimics the Dnmt3a-Dnmt3L heterodimer, which are colored in *pink* and *wheat*, respectively

to that of the NtDRM MTase domain. As a result, the nearly identical folding topology between NtDRM and other type I DNA MTases indicates both a common catalytic mechanism and a common evolutionary ancestor.

The NtDRM MTase domain forms a homodimeric arrangement as observed both in solution and from packing alignments in the crystal (Fig. 3e). The dimer interface is composed of a hydrophobic core together with several pairs of intermolecular salt bridges (Zhong et al. 2014). It is interesting that the NtDRM homodimer interface mimics the Dnmt3a-Dnmt3L heterodimer interface (Fig. 3e) (Jia et al. 2007; Zhong et al. 2014). Although *Arabidopsis* has an inactive DRM homolog DRM3, which is essential for the establishment of DNA methylation by RdDM (Henderson et al. 2010), DRM3 appears not to have obvious interactions with DRM2, given that DRM3 cannot be detected in immunoprecipitation followed by mass spectrometry assays using epitope-tagged DRM2 (Zhong et al. 2015). In the *drm3* strain, Pol V occupancy extends to some more additional loci but the Pol V transcripts abundancy is reduced, revealing a potential role of DRM3 in the stabilization of Pol V and/or mediating Pol V transcript elongation (Zhong et al. 2015). Thus, DRM3 in plants appears not to function like Dnmt3L by acting as cofactor of the active DNA MTase but likely functions in regulating RdDM by association with Pol V (Zhong et al. 2015). The disruption of the dimerization interface by five mutations of NtDRM results in disruption of its enzymatic activity, indicating that dimerization is biochemically essential (Zhong et al. 2014). A plausible explanation of the homodimerization of DRM MTase domains is that dimerization could help to maintain the conformation of the catalytic loop, because the C-terminal portion of the catalytic loop is involved in dimerization interface formation. The mammalian de novo DNA MTase requires a Dnmt3L-Dnmt3a-Dnmt3a-Dnmt3L tetrameric arrangement (Jia et al. 2007). Although the DRM homodimer mimics the Dnmt3a-Dnmt3L dimer interface, there is no interface in DRM that corresponds to the Dnmt3a-Dnmt3a interface. The Dnmt3a-Dnmt3a interface is essential for its activity and is believed to double the DNA-binding surface so as to overcome the shortage of the DNA-binding surface, because of the small TRD subdomain of Dnmt3a (Jia et al. 2007; Jurkowska et al. 2008). In DRM, the TRD subdomain is bigger than that of Dnmt3a and can form a continuous large negatively charged surface, which is capable of accessing the DNA substrate (Zhong et al. 2014). Thus, the Dnmt3a-Dnmt3a like interface appears not to be required in the DRM case.

The UBA domains of DRM have been shown to be critical for its in vivo but not in vitro DNA MTase activity, revealing an important regulatory but not catalytic role (Henderson et al. 2010; Zhong et al. 2014). So far, it is still unclear how the UBA domains participate in DRM function. A report revealed that the UBA domains of rice OsDRM2 can interact with the ATP-dependent RNA helicase eIF4A, which links RNA higher structure to DRM (Dangwal et al. 2013). Another plausible connection might be in the recognition of some ubiquitination modifications by the UBA domains, which is a common function of UBA domains. However, all these hypotheses still need further testing and eventual validation.

2.4 Potential Mechanism of MET1 in CG Methylation Maintenance

Unlike the extensive cross talk between non-CG methylation and H3K9me2, CG methylation seems independent of H3K9me2 (Du et al. 2015). In the *kyp suvh5 suvh6* triple mutant, which eliminates most of the H3K9me2 mark, only a limited reduction is observed in CG methylation (Stroud et al. 2013). Once established, CG methylation is subsequently faithfully maintained by MET1 (Kankel et al. 2003). MET1 is an ortholog of mammalian Dnmt1 with several distinct features in its sequence. By comparing the sequence motifs of mouse Dnmt1, whose structure has been extensively studied, and *Arabidopsis* MET1 (Fig. 4a), it becomes apparent that the two C-terminal BAH domains (BAH1 and BAH2) and the DNA MTase domain portion exhibit similarities with a sequence identity of 36 %. The high sequence homology between the catalytic portion of MET1 and Dnmt1 suggests that MET1 may share a similar type I DNA MTase domain fold and a common catalytic mechanism as Dnmt1. The BAH2 domain of Dnmt1 has a long loop extending outward whose tip contacts the TRD subdomain of the MTase fold (Fig. 4b) (Song et al. 2011), suggestive of a plausible regulatory role for substrate DNA binding through adjustment of the TRD position by the BAH2 domain. However, neither the interacting region on the TRD (highlighted in red in Fig. 4b) nor the BAH2 loop (highlighted in blue in Fig. 4b) of Dnmt1 is conserved in MET1 (Figs. 4c, d), indicating that MET1 may lack this type of BAH domain-mediated regulation of the TRD subdomain. Toward the N-terminus, Dnmt1 has a CXXC domain (highlighted in yellow in Fig. 4b), which can specifically recognize unmethylated CG sites and subsequently position a loop (highlighted in green in Fig. 4b) between the DNA and the active site of Dnmt1 to block the MTase activity (Song et al. 2011). This auto-inhibition mechanism can protect Dnmt1's function as a maintenance MTase by insuring cytosine methylation of the daughter strand on hemi-methylated CG DNA but not on unmethylated CG DNA, which can be captured by the CXXC domain (Song et al. 2011). However, this type of auto-inhibitory effect does not biochemically show in the full-length protein, revealing that there are domain rearrangements which are dependent also on all other domains as illustrated by the different effect of the CXXC domain observed with truncated and full-length Dnmt1 (Bashtrykov et al. 2012; Song et al. 2011). In contrast, plant MET1 does not have a CXXC domain (Fig. 4a), suggesting that such an auto-inhibitory mechanism is not conserved in plants. In addition, Dnmt1 contains a replication foci targeting domain (RFTD), which has been shown to interact with the catalytic domain in the absence of DNA, thereby achieving an additional layer of auto-inhibition (Takeshita et al. 2011; Bashtrykov et al. 2014; Syeda et al. 2011). In contrast, plant MET1 has two putative RFTDs as predicted by the Pfam server (Finn et al. 2014) (Fig. 4a), but their exact function remains unclear. Further structural and functional studies may shed light on the function and regulation of the plant MET1 DNA MTase.

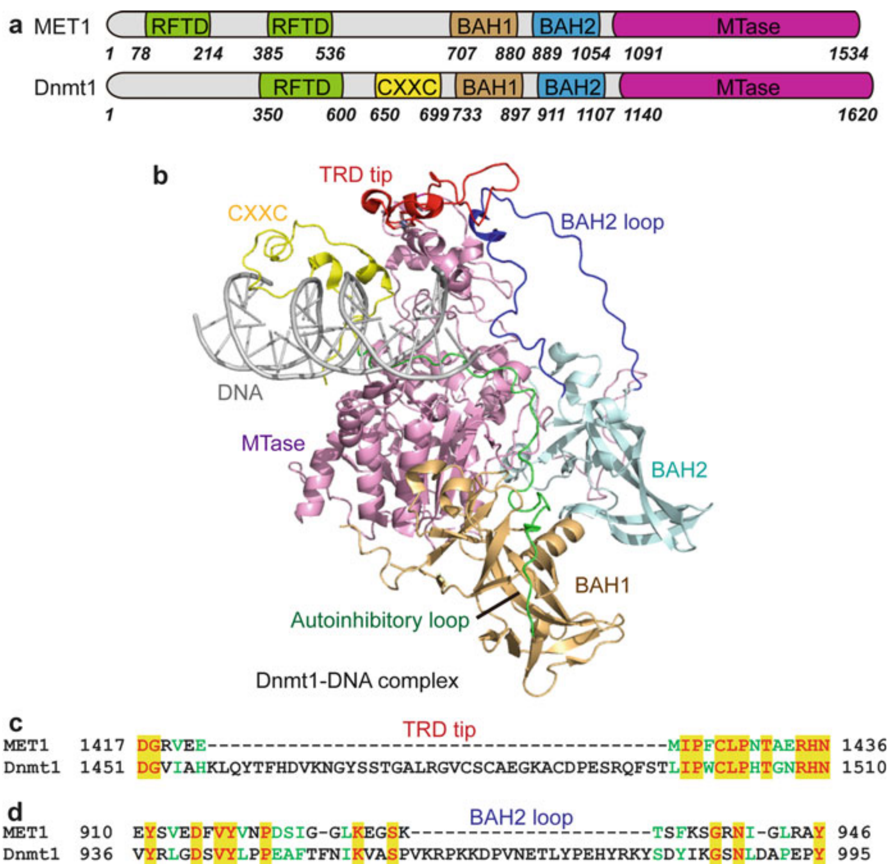


Fig. 4 Potential mechanistic insights into MET1 function. (**a**) Comparison between the domain architectures of *Arabidopsis* MET1 and its ortholog mouse Dnmt1. MET1 contains one additional putative RFTD but lack the CXXC domain. (**b**) Structure of Dnmt1-DNA complex in an auto-inhibition mode (PDB code: 3PT6). The CXXC domain, auto-inhibition loop, BAH1, BAH2, MTase domains, and the bounded DNA are colored in *yellow*, *green*, *orange*, *light cyan*, *light magenta*, and *silver*, respectively. A tip within the loop of the BAH2 domain can interact with the TRD subdomain of the MTase to form a novel regulatory mechanism, with the interaction highlighted in blue and red, respectively. (**c**) Sequence alignment of the TRD subdomain of MTase between Dnmt1 and MET1 indicates that MET1 lack the interaction region within the TRD. Strictly conserved residues are highlighted in red with yellow background and the moderately conserved residues are colored in *green*. (**d**) Sequence alignment of the BAH2 loop region of Dnmt1 and MET1 showing that MET1 lacks the BAH2 loop. Therefore, MET1 does not have the BAH2-TRD interaction and this type of enzymatic regulation

3 Conclusion and Perspective

In plants, DNA methylation functions importantly in the suppression of TE and repeat sequences so as to act as a genomic immune response to overcome the abundantly distributed TEs across plant genomes (Kim and Zilberman 2014). In contrast

to the mammalian DNA methylation system, plant DNA methylation is both more diversified and more complex and includes methylation of cytosines in all three sequence (CG, CHG, and CHH) contexts. Notably, there are four active DNA MTases (DRM2, MET1, CMT3, CMT2), in which the DRM and CMT3/2 are plant-specific DNA MTases, while MET1 has similarities with mammalian Dnmt1. In general, all the available structures of the catalytic domains of known plant DNA MTases adopt typical type I DNA MTase folds (Zhong et al. 2014; Du et al. 2012), sharing the same catalytic mechanism as mammalian DNA MTases such as Dnmt3a and Dnmt1 (Jia et al. 2007; Song et al. 2011, 2012). However, there is diversity in the regulation of these DNA MTase, while being enriched with their own additional features. The CMT family MTases can be regulated through their BAH and chromo domains by recognition of H3K9me2 (Stroud et al. 2014; Du et al. 2012). The RdDM-driven DRM2 function is regulated by H3K9me2 through SHH1 and by pre-methylated DNA through SUVH2/9 (Johnson et al. 2014; Law et al. 2013). The UBA domains of DRM may also play plausible regulatory roles through so far an unknown pathway. MET1 may be regulated similarly to Dnmt1, but with its own features because it lacks both the CXXC domain and the regulatory loop within the BAH2 domain. The current structural studies on plant DNA MTases have established regulatory mechanisms for CMT and DRM. Further studies on the UBA domains of DRM and MET1 may resolve additional details about the regulation of RdDM and the maintenance of CG methylation in plants. Moreover, structures of plant DNA MTase in complex with substrate DNA are required so as to reveal the molecular mechanisms underlying the observed sequence specificity associated with each plant DNA MTase.

It worth to note that in addition to 5mC, the 6mA mark was recently reported to be located at transcription start sites that mark the location of active genes in green algae *Chlamydomonas reinhardtii* (Fu et al. 2015). In addition to the algae, 6mA was also detected in higher plants (Vaniushin et al. 1971). In particular, it was found that the *Arabidopsis DRM2* gene locus contains 6mA in some GATC sequence contexts (Ashapkin et al. 2002). However, the functional role of 6mA in higher plants is still controversial and requires additional investigation. The enzymes responsible for 6mA generation and elimination in plants have not yet been identified. Thus, the further investigation of 6mA in higher plants may open a new window for the plant DNA methylation studies.

Acknowledgment I apologize to those whose work was not discussed due to space limitation. I would like to thank Dr. Steven E. Jacobsen, Dr. Suhua Feng (University of California, Los Angeles), and Dr. Dinshaw J. Patel (Memorial Sloan-Kettering Cancer Center) for critical reading of the manuscript and helpful discussions. This work was supported by the Ministry of Science and Technology of China (2016YFA0503200), the Thousand Young Talents Program of China, and the Chinese Academy of Sciences.

References

Arita K, Ariyoshi M, Tochio H, Nakamura Y, Shirakawa M. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature. 2008;455(7214):818–21. doi:10.1038/nature07249.

Ashapkin VV, Kutueva LI, Vanyushin BF. The gene for domains rearranged methyltransferase (DRM2) in Arabidopsis thaliana plants is methylated at both cytosine and adenine residues. FEBS Lett. 2002;532(3):367–72.

Avvakumov GV, Walker JR, Xue S, Li Y, Duan S, Bronner C, et al. Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. Nature. 2008;455(7214):822–5. doi:10.1038/nature07273.

Bashtrykov P, Jankevicius G, Smarandache A, Jurkowska RZ, Ragozin S, Jeltsch A. Specificity of Dnmt1 for methylation of hemimethylated CpG sites resides in its catalytic domain. Chem Biol. 2012;19(5):572–8. doi:10.1016/j.chembiol.2012.03.010.

Bashtrykov P, Rajavelu A, Hackner B, Ragozin S, Carell T, Jeltsch A. Targeted mutagenesis results in an activation of DNA methyltransferase 1 and confirms an autoinhibitory role of its RFTS domain. Chembiochem. 2014;15(5):743–8. doi:10.1002/cbic.201300740.

Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE. Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. PLoS One. 2008;3(9):e3156. doi:10.1371/journal.pone.0003156.

Blus BJ, Wiggins K, Khorasanizadeh S. Epigenetic virtues of chromodomains. Crit Rev Biochem Mol Biol. 2011;46(6):507–26. doi:10.3109/10409238.2011.619164.

Cao X, Jacobsen SE. Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. Proc Natl Acad Sci U S A. 2002a;99 Suppl 4:16491–8. doi:10.1073/pnas.162371599.

Cao X, Jacobsen SE. Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. Curr Biol. 2002b;12(13):1138–44.

Castel SE, Martienssen RA. RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. Nat Rev Genet. 2013;14(2):100–12. doi:10.1038/nrg3355.

Cheng X, Kumar S, Posfai J, Pflugrath JW, Roberts RJ. Crystal structure of the HhaI DNA methyltransferase complexed with S-adenosyl-L-methionine. Cell. 1993;74(2):299–307.

Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, et al. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature. 2008;452(7184):215–9. doi:10.1038/nature06745.

Dangwal M, Malik G, Kapoor S, Kapoor M. De novo methyltransferase, OsDRM2, interacts with the ATP-dependent RNA helicase, OseIF4A, in rice. J Mol Biol. 2013;425(16):2853–66. doi:10.1016/j.jmb.2013.05.021.

Du J, Johnson LM, Groth M, Feng S, Hale CJ, Li S, et al. Mechanism of DNA methylation-directed histone methylation by KRYPTONITE. Mol Cell. 2014;55(3):495–504. doi:10.1016/j.molcel.2014.06.009.

Du J, Johnson LM, Jacobsen SE, Patel DJ. DNA methylation pathways and their crosstalk with histone methylation. Nat Rev Mol Cell Biol. 2015;16(9):519–32. doi:10.1038/nrm4043.

Du J, Zhong X, Bernatavichute YV, Stroud H, Feng S, Caro E, et al. Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell. 2012;151(1):167–80. doi:10.1016/j.cell.2012.07.034.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. Nucleic Acids Res. 2014;42(Database issue):D222–30. doi:10.1093/nar/gkt1223.

Finnegan EJ, Kovac KA. Plant DNA methyltransferases. Plant Mol Biol. 2000;43(2–3):189–201.

Fu Y, Luo GZ, Chen K, Deng X, Yu M, Han D, et al. N(6)-methyldeoxyadenosine marks active transcription start sites in chlamydomonas. Cell. 2015;161(4):879–92. doi:10.1016/j.cell.2015.04.010.

Goll MG, Bestor TH. Eukaryotic cytosine methyltransferases. Annu Rev Biochem. 2005;74:481–514. doi:10.1146/annurev.biochem.74.010904.153721.

Haag JR, Pikaard CS. Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. Nat Rev Mol Cell Biol. 2011;12(8):483–92. doi:10.1038/nrm3152.

Haag JR, Ream TS, Marasco M, Nicora CD, Norbeck AD, Pasa-Tolic L, et al. In vitro transcription activities of Pol IV, Pol V, and RDR2 reveal coupling of Pol IV and RDR2 for dsRNA synthesis in plant RNA silencing. Mol Cell. 2012;48(5):811–8. doi:10.1016/j.molcel.2012.09.027.

Hashimoto H, Horton JR, Zhang X, Bostick M, Jacobsen SE, Cheng X. The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. Nature. 2008;455(7214):826–9. doi:10.1038/nature07280.

He XJ, Ma ZY, Liu ZW. Non-coding RNA transcription and RNA-directed DNA methylation in Arabidopsis. Mol Plant. 2014;7(9):1406–14. doi:10.1093/mp/ssu075.

Henderson IR, Deleris A, Wong W, Zhong X, Chin HG, Horwitz GA, et al. The de novo cytosine methyltransferase DRM2 requires intact UBA domains and a catalytically mutated paralog DRM3 during RNA-directed DNA methylation in Arabidopsis thaliana. PLoS Genet. 2010;6(10):e1001182. doi:10.1371/journal.pgen.1001182.

Jackson JP, Johnson L, Jasencakova Z, Zhang X, PerezBurgos L, Singh PB, et al. Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in Arabidopsis thaliana. Chromosoma. 2004;112(6):308–15. doi:10.1007/s00412-004-0275-7.

Jackson JP, Lindroth AM, Cao X, Jacobsen SE. Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature. 2002;416(6880):556–60. doi:10.1038/nature731.

Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature. 2007;449(7159):248–51. doi:10.1038/nature06146.

Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, Callis J, et al. The SRA methyl-cytosine-binding domain links DNA and histone methylation. Curr Biol. 2007;17(4):379–84. doi:10.1016/j.cub.2007.01.009.

Johnson LM, Du J, Hale CJ, Bischof S, Feng S, Chodavarapu RK, et al. SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. Nature. 2014;507(7490):124–8. doi:10.1038/nature12931.

Johnson LM, Law JA, Khattar A, Henderson IR, Jacobsen SE. SRA-domain proteins required for DRM2-mediated de novo DNA methylation. PLoS Genet. 2008;4(11):e1000280. doi:10.1371/journal.pgen.1000280.

Jurkowska RZ, Anspach N, Urbanke C, Jia D, Reinhardt R, Nellen W, et al. Formation of nucleoprotein filaments by mammalian DNA methyltransferase Dnmt3a in complex with regulator Dnmt3L. Nucleic Acids Res. 2008;36(21):6656–63. doi:10.1093/nar/gkn747.

Kankel MW, Ramsey DE, Stokes TL, Flowers SK, Haag JR, Jeddeloh JA, et al. Arabidopsis MET1 cytosine methyltransferase mutants. Genetics. 2003;163(3):1109–22.

Kim MY, Zilberman D. DNA methylation as a system of plant genomic immunity. Trends Plant Sci. 2014;19(5):320–6. doi:10.1016/j.tplants.2014.01.014.

Kuo AJ, Song J, Cheung P, Ishibe-Murakami S, Yamazoe S, Chen JK, et al. The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. Nature. 2012;484(7392):115–9. doi:10.1038/nature10956.

Law JA, Du J, Hale CJ, Feng S, Krajewski K, Palanca AM, et al. Polymerase IV occupancy at RNA-directed DNA methylation sites requires SHH1. Nature. 2013;498(7454):385–9. doi:10.1038/nature12178.

Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet. 2010;11(3):204–20. doi:10.1038/nrg2719.

Lindroth AM, Cao X, Jackson JP, Zilberman D, McCallum CM, Henikoff S, et al. Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. Science. 2001;292(5524):2077–80. doi:10.1126/science.1059745.

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, et al. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell. 2008;133(3):523–36. doi:10.1016/j.cell.2008.03.029.

Liu ZW, Shao CR, Zhang CJ, Zhou JX, Zhang SW, Li L, et al. The SET domain proteins SUVH2 and SUVH9 are required for Pol V occupancy at RNA-directed DNA methylation loci. PLoS Genet. 2014;10(1):e1003948. doi:10.1371/journal.pgen.1003948.

Matzke MA, Mosher RA. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat Rev Genet. 2014;15(6):394–408. doi:10.1038/nrg3683.

Patel DJ, Wang Z. Readout of epigenetic modifications. Annu Rev Biochem. 2013;82:81–118. doi:10.1146/annurev-biochem-072711-165700.

Rajakumara E, Law JA, Simanshu DK, Voigt P, Johnson LM, Reinberg D, et al. A dual flip-out mechanism for 5mC recognition by the Arabidopsis SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation in vivo. Genes Dev. 2011;25(2):137–52. doi:10.1101/gad.1980311.

Shen X, De Jonge J, Forsberg SK, Pettersson ME, Sheng Z, Hennig L, et al. Natural CMT2 variation is associated with genome-wide methylation changes and temperature seasonality. PLoS Genet. 2014;10(12):e1004842. doi:10.1371/journal.pgen.1004842.

Song J, Rechkoblit O, Bestor TH, Patel DJ. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. Science. 2011;331(6020):1036–40. doi:10.1126/science.1195380.

Song J, Teplova M, Ishibe-Murakami S, Patel DJ. Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. Science. 2012;335(6069):709–12. doi:10.1126/science.1214453.

Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, et al. Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nat Struct Mol Biol. 2014;21(1):64–72. doi:10.1038/nsmb.2735.

Stroud H, Greenberg MV, Feng S, Bernatavichute YV, Jacobsen SE. Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. Cell. 2013;152(1–2):352–64. doi:10.1016/j.cell.2012.10.054.

Syeda F, Fagan RL, Wean M, Avvakumov GV, Walker JR, Xue S, et al. The replication focus targeting sequence (RFTS) domain is a DNA-competitive inhibitor of Dnmt1. J Biol Chem. 2011;286(17):15344–51. doi:10.1074/jbc.M110.209882.

Takeshita K, Suetake I, Yamashita E, Suga M, Narita H, Nakagawa A, et al. Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). Proc Natl Acad Sci U S A. 2011;108(22):9055–9. doi:10.1073/pnas.1019629108.

Vaniushin BF, Kadyrova D, Karimov K, Belozerskii AN. [Minor bases in DNA of higher plants]. Biokhimiia. 1971;36(6):1251–8.

Vanyushin BF, Ashapkin VV. DNA methylation in higher plants: past, present and future. Biochim Biophys Acta. 2011;1809(8):360–8. doi:10.1016/j.bbagrm.2011.04.006.

Wada Y, Ohya H, Yamaguchi Y, Koizumi N, Sano H. Preferential de novo methylation of cytosine residues in non-CpG sequences by a domains rearranged DNA methyltransferase from tobacco plants. J Biol Chem. 2003;278(43):42386–93. doi:10.1074/jbc.M303892200.

Woo HR, Dittmer TA, Richards EJ. Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in Arabidopsis. PLoS Genet. 2008;4(8):e1000156. doi:10.1371/journal.pgen.1000156.

Yang N, Xu RM. Structure and function of the BAH domain in chromatin biology. Crit Rev Biochem Mol Biol. 2013;48(3):211–21. doi:10.3109/10409238.2012.742035.

Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, et al. The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell. 2013;153(1):193–205. doi:10.1016/j.cell.2013.02.033.

Zhang H, Ma ZY, Zeng L, Tanaka K, Zhang CJ, Ma J, et al. DTF1 is a core component of RNA-directed DNA methylation and may assist in the recruitment of Pol IV. Proc Natl Acad Sci U S A. 2013;110(20):8290–5. doi:10.1073/pnas.1300585110.

Zhao Y, Chen X. Noncoding RNAs and DNA methylation in plants. Natl Sci Rev. 2014;1(2):219–29. doi:10.1093/nsr/nwu003.

Zhong X, Du J, Hale CJ, Gallego-Bartolome J, Feng S, Vashisht AA, et al. Molecular mechanism of action of plant DRM de novo DNA methyltransferases. Cell. 2014;157(5):1050–60. doi:10.1016/j.cell.2014.03.056.

Zhong X, Hale CJ, Nguyen M, Ausin I, Groth M, Hetzel J, et al. Domains rearranged methyltransferase3 controls DNA methylation and regulates RNA polymerase V transcript abundance in Arabidopsis. Proc Natl Acad Sci U S A. 2015;112(3):911–6. doi:10.1073/pnas.1423603112.

DNA Methylation and Gene Regulation in Honeybees: From Genome-Wide Analyses to Obligatory Epialleles

Laura Wedd and Ryszard Maleszka

Abstract
In contrast to heavily methylated mammalian genomes, invertebrate genomes are only sparsely methylated in a 'mosaic' fashion with the majority of methylated CpG dinucleotides found across gene bodies. Importantly, this gene body methylation is frequently associated with active transcription, and studies in the honeybee have shown that there are strong links between gene body methylation and alternative splicing. Additional work also highlights that obligatory methylated epialleles influence transcriptional changes in a context-specific manner. Here we discuss the current knowledge in this emerging field and highlight both similarities and differences between DNA methylation systems in mammals and invertebrates. Finally, we argue that the relationship between genetic variation, differential DNA methylation, other epigenetic modifications and the transcriptome must be further explored to fully understand the role of DNA methylation in converting genomic sequences into phenotypes.

Abbreviations

5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
ALK	Anaplastic lymphoma kinase
CpG	Cytosine and guanine dinucleotide separated by one phosphate in DNA
CTCF	CCCTC-binding protein
DBP	DNA-binding protein

L. Wedd • R. Maleszka (✉)
Research School of Biology, The Australian National University,
Canberra ACT 2601, Australia
e-mail: ryszard.maleszka@anu.edu.au

© Springer International Publishing Switzerland 2016

A. Jeltsch, R.Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*,
Advances in Experimental Medicine and Biology 945,
DOI 10.1007/978-3-319-43624-1_9

DNMT	DNA methyltransferase
LAM	Lysosomal alpha-mannosidase
MeCP	Methyl-CpG-binding factor
miRNA	microRNA small non-coding RNA of about 22 nucleotides
mRNA	messenger RNA
PTM	Post-translational modification
RdDM	RNA-directed DNA methylation system
TET	Ten-eleven translocation enzyme

1 Introduction

Amongst all fields of biomedical research, it is ‘epigenetics’ that is emerging as a principal discipline bridging the gap between genotype and phenotype. One reason for such a prominent standing of epigenetic concepts in modern biology is the flexible and ever-expanding definition of what the term ‘epigenetics’ actually means. Over 70 years ago, Waddington originally introduced the word ‘epigenetics’ (derived from the word ‘epigenesis’) as ‘a suitable name for the branch of biology which studies the causal interactions between genes and their products which bring phenotype into being’ (Jablonka and Lamb 2002). This new definition implied that translating the genetic blueprint into a functional organism requires a control system whose mode of action is over and above, or in addition to, the classical genotype. His ideas were incited by the realisation that phenotypes are remarkably stable in spite of environmental pressure, a phenomenon that he referred to as developmental canalisation, whereby development is buffered against environmental or genetic variation by evolutionarily selected gene networks. He illustrated this concept by the famous epigenetic landscape with valleys or channels representing the optimal developmental trajectories. The key idea behind this imaginary topography is that a set of instructions carving the trajectory towards an optimal phenotype is selected by responses to recurring environmental insults in an organism’s adaptive niche (Waddington 1957). Phenotypic plasticity is another feature of the living world, distinct from canalisation but equally important, that also is independent of the underlying DNA sequence. Cellular differentiation in multicellular organisms or phenotypic polymorphisms in social insects are generated by epigenetic mechanisms, not by genetic differences. These two contrasting aspects of phenotypic changeability that are uncoupled from genetic variation are central to the study of epigenetics. At present, epigenetics is a very wide field of study, covering virtually all aspects of biology ranging from morphogenesis to transgenerational epigenetic inheritance. For more detailed discussions on the origins and changing concept of epigenetics, see recent review and opinion articles (Jablonka and Lamm 2012; Haig 2004, 2012).

In modern times the distinction between genetic and epigenetic control system has been associated with specific biochemical processes. Typically, DNA methylation and histone modifications are itemised as epigenetic mechanisms, but non-coding RNAs, chromatin remodelling and even prions are often included (Halfmann and Lindquist 2010). In this context, it is worth noting that only DNA methylation

is part of the DNA molecule chemical entity. It is thus clearly distinguished from other epigenetic mechanisms, such as chromatin modification and non-coding RNAs, which are associated with, but separate from, DNA. For over 50 years DNA methylation, or more specifically cytosine methylation, has been studied extensively in mammals, initially as a mechanism of gene silencing via hyper-methylation of promoters associated with the CpG islands and later as a genome-wide modification. In contrast, only a couple of studies in the late 1990s have shown that 5-methylcytosine can be detected in DNAs extracted from various invertebrate species (Tweedie et al. 1997; Regev et al. 1998), and it is largely through the recent honeybee genomic research that methylomics took a centre stage as an important mechanism of gene regulation in insects and invertebrates. In this article, we discuss DNA methylation as part of gene regulatory system in invertebrates, with special reference to the honeybee, *Apis mellifera*.

2 Genotype to Phenotype

The existence of a multicellular organism depends upon the transformation of an apparently simple, static genetic ‘code’ into function. With perfect timing the genetic information contained within that first single cell is translated into a series of complex cellular signals that guide development. As each cell differentiates, unique transcriptional profiles are established and their functional roles are specified. This transition from genotype to phenotype, termed *epigenesis*, is infinitely complex and results from interactions between the underlying genetic sequence, chemical modifications on DNA and chromatin and environmental cues. Yet, how exactly do these layers of information contribute to establishing phenotype?

Foremost, the DNA sequence itself interacts with the transcriptional machinery to produce the multitude of proteins necessary for a functioning organism. The human genome, for instance, encodes approximately 21,000 protein-coding genes whose sequences are transcribed to generate our proteome (Clamp et al. 2007). Gene transcription occurs when RNA polymerase II (RNA Pol II) interacts with these genomic regions to produce a transcript. This is governed by numerous functional elements, such as enhancers or promoters, associated with the coding regions that interact with various activators and transcription factors to facilitate the assembly of the pre-initiation complex and subsequent transcription by RNA Pol II (Fig. 1).

Importantly, to generate the full repertoire of proteins required to build a complex and fully functioning organism, each gene must be expressed in a precise spatio-temporal pattern. To achieve this, numerous control mechanisms have evolved to tightly regulate transcription. These range from modulating transcription initiation and elongation, which can alter whether a gene is expressed or not and how a transcript is processed, to post-translational processes that can, for instance, fine-tune the level of a transcript through the degradation of an mRNA product. In addition, epigenetic control systems have evolved to coordinate the action of thousands of genes and to provide an interface between the genome and environment. Each of these layers represents an important mechanism through which gene expression can be controlled.

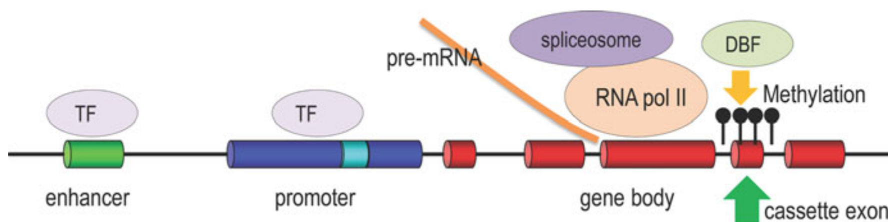


Fig. 1 General model of transcription in the honeybee showing the key elements involved in the initiation and elongation of pre-mRNA. Splicing is assumed to occur co-transcriptionally with DNA methylation affecting the conditional usage of weak exons. In insects, methylated cytosines have been found predominantly in intragenic regions, often in proximity of splice sites. DNA methylation may affect binding kinetics of sequence-specific factors (DBP) either by disrupting their binding, as in the CTCF model (Shukla et al. 2011), or by recruiting methyl-CpG-binding factors such as MeCP2 (Maunakea et al. 2013). In both models, such factors promote exon recognition either by pausing the spliceosome/RNA pol II complex (Shukla et al. 2011) or by recruiting histone deacetylases to maintain low acetylation levels of alternatively spliced exons, which could reduce transcription elongation (Maunakea et al. 2013). TF- transcription factor

3 The Epigenetic Control of Gene Expression

Epigenetics encompasses those mechanisms and processes that are involved in facilitating transcriptional changes via various covalent modifications made to DNA itself or the histone proteins around which DNA is wrapped; these post-translational modifications (PTMs) are numerous and include methylation, acetylation, phosphorylation, ubiquitination and many other PTMs. The manner in which these modifications can influence transcription is complex, and it is likely that a given mark can lead to both gene activation and repression, depending on its location and the genomic context (Kouzarides 2007; Jones 2012). Importantly, none of these modifications works in isolation, instead they interact together to produce a unique epigenetic cellular 'signature'.

This signature, termed a cell's *epigenome*, describes all epigenetic modifications found across the genome. For each cell type within an organism, and across different developmental stages and disease states, the epigenome will vary, leading to a vast number of possible epigenomes (Bernstein et al. 2007). Through analysing patterns of DNA methylation, histone modifications, chromatin accessibility and RNA expression across diverse cell lineages, human epigenomic maps are being catalogued (Kundaje et al. 2015; Bernstein et al. 2010). These large-scale pursuits are leading to a greater understanding of how the epigenome contributes to cell specification and development and how alterations to the epigenome contribute to disease and phenotypic variation. However, defining the exact role that a given epigenetic modification plays in directing transcriptional changes still remains a challenge. This difficulty will be discussed here in the context of one of the extensively studied epigenomic modifications, DNA methylation.

4 DNA Methylation

In all vertebrates and in the majority of invertebrates, the cytosine nucleotide in DNA can be modified by the addition of a methyl group to its 5-carbon. This modification typically occurs at cytosines in CpG dinucleotides and is prevalent in mammalian genomes where up to 80 % (over 20 million) (Smith and Meissner 2013) of cytosines in the CpG context are methylated (Bird 2002). In comparison, the proportion of methylated CpGs in insect genomes is an order of magnitude lower ranging from approximately 0.1 to 1 % (Lyko and Maleszka 2011). The realisation that this covalent modification could stably propagate information throughout development led to considerable research into its role as an epigenetic mark (Holliday and Pugh 1975; Riggs 1975). It has since been shown that DNA methylation plays a key role in biological processes such as X-inactivation, genomic imprinting and transposon silencing across a number of organisms. In each of these instances, methylation is engaged in gene silencing, and this remains one of the more commonly considered functions of DNA methylation. This is evident in mammals where the methylation of promoter regions has long since been associated with gene repression (Jones and Takai 2001). However, considerable variation in the distribution of methylation across organisms and additional evidence that methylation associates with active transcription indicates that the relationship between DNA methylation and transcription is complex and not fully understood. Insight into its functionality can be gained through understanding how DNA methylation patterns are established and by analysing the evolutionary origins of this mark and its genomic location (Suzuki and Bird 2008; Regev et al. 1998).

Box: Epigenetics and Epigenomics

- *Epigenetics* is a broad term that covers many areas of biological research that involve gene regulation. It includes both heritable and non-heritable chemical modifications, such as DNA methylation and hydroxymethylation, histone modifications, microRNAs and non-coding RNAs.
- *Epigenetic processes* can be defined as control systems that combine a number of epigenomic modifications whose actions, responsive to both internal and external environments, underlie the self-organising properties of gene networks.
- *Epigenomics* is a study of genome-wide modifications of DNA and chromatin that allows mapping the epigenetic status of all genes in a context-dependent manner, e.g., in various tissues or throughout development. Each organism has hundreds of epigenomes, minimally one for each cell type.
- *The epigenetic status* of a gene can be altered through environmental events such as changes in diet or exposure to toxins often resulting in disease.

Table. 1 Examples of mosaic distribution of DNA methylation toolkit in selected metazoan species

Lineage	Species	DNMT 1	DNMT 3	TET	Methyl CpGs
Mammalia	All	●	●●●	●●●	Yes
Hymenoptera	*Apis mellifera* (honeybee)	●●	●	●	Yes
	Nasonia spp. (parasitic wasps)	●●●	●	●	Yes
Homoptera	*Acyrthosiphon pisum* (pea aphid)	●●	●	●	Yes
Coleoptera	*Tribolium castaneum* (flour beetle)	●	–	●	Yes
Lepidoptera	*Bombyx mori* *Spodoptera spp.*	●	–	●	Yes
Diptera	Flies, mosquitos	–	–	●[a]	No
Nematoda	*Caenorhabditis elegans* (free living roundworm)	–	–	–	No
	Trichuris trichiura (whipworm, parasitic)	–	●[b]	–	?
Placozoa	*Trichoplax adhaerens* (the only extant member of this phylum)	–	–	●	No
Cnidaria	*Nematostella vectensis* (sea anemone)	●	●	●	Yes

Black dots indicate the number of genes encoding DNMT paralogs in various species

[a]In *Drosophila* TET demethylates N^6-methyladenine (6 mA) and this process correlates with transposon expression (Zhang et al. 2015)

[b]A highly conserved relative of DNMT3 in *T. trichiura* available in GenBank (KFD71641.1) needs to be validated by additional analyses. If confirmed, the whipworm would be the only nematode with a gene encoding a DNMT and the only species so far having DNMT3, but no DNMT1 and TET

4.1 Establishing DNA Methylation Patterns in Invertebrates

The overall level and patterns of DNA methylation that are established across a genome, what is termed a *methylome*, are determined in a cell- and tissue-specific manner. Throughout embryogenesis and cell differentiation, DNA methyltransferases (DNMTs) establish DNA methylation patterns and then maintain these patterns across cell divisions. Two families of DNMTs exist: one referred to as DNMT1, which preferentially methylates hemimethylated cytosine residues and the other as DNMT3 family, which catalyses the de novo methylation of DNA (Goll and Bestor 2005). Another important protein family relevant for DNA methylation are TET dioxygenases that remove methyl groups from modified cytosines (see a section below). All these proteins are well conserved and they are present across a wide variety of vertebrates and invertebrates. However, as shown in Table 1, one salient aspect of DNA methylation is its dispensability. The distribution of the DNA methylation toolkit in *Metazoa* is mosaic with a variety of distinct patterns ranging from

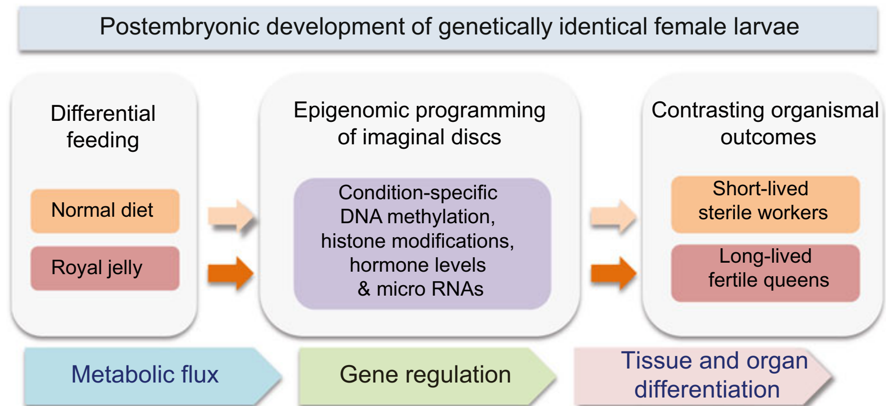


Fig. 2 Outline of postembryonic developmental reprogramming in honeybees by nutritionally driven epigenomic mechanisms

a total loss of both DNMT and TET enzymes to partial losses and duplications. Many species, including most nematodes, advanced Dipteran insects and *Placozoa* have lost DNA methylating enzymes and apparently recruited other epigenetic mechanisms to regulate gene expression. Given this contrasting utilisation of an ancient biochemical modification, the extent to which DNA methylation is advantageous if maintained in a given lineage remains unclear (Lyko and Maleszka 2011; Kucharski et al. 2008; Zemach et al. 2010).

Initially, it was considered that DNA methylation patterns were established by the de novo activity of the DNMT3 family during embryogenesis and then these patterns were maintained by DNMT1, the 'maintenance' DNMT. However, recent work indicates that the classical roles of DNMT1 and DNMT3s in establishing methylation patterns need to be redefined to include the apparent de novo activity of DNMT1 and DNMT3s' contribution to maintenance (Jeltsch and Jurkowska 2014). The variation of DNMTs across invertebrates is also suggestive of diverse roles for these enzymes; *Bombyx mori*, for instance, lacks DNMT3 but still methylates its genome and it is therefore likely that DNMT1 provides de novo activity in this organism (Lyko and Maleszka 2011; Xiang et al. 2010). The honeybee, *A. mellifera*, is known to possess two copies of DNMT1 together with one DNMT3 gene and has become a focal model for exploring the relationship between DNA methylation and phenotypic plasticity in invertebrates. For instance, differential methylation patterns in this organism are correlated with the establishment of two distinct castes, that of queen and worker, from identical genetic information (Lyko and Maleszka 2011; Kucharski et al. 2008). This process is driven by differential feeding during postembryonic development that affects metabolic fluxes, hormonal changes, global DNA methylation and gene expression (Maleszka 2008; Miklos and Maleszka 2011; Foret et al. 2012; Kucharski et al. 2015) (Fig. 2). Newly hatched female larvae fed a complex diet known as royal jelly develop into long-lived reproductive queens, whereas larvae fed less nutritious worker jelly develop into functionally sterile

short-lived workers. Interference with DNA methylation by knocking down DNMT3 in larvae reared in vitro mimics the effect of royal jelly on postembryonic development suggesting that dietary ingredients have the capacity to affect epigenomic settings (Kucharski et al. 2008). Although this result clearly implicates DNA methylation in a diet-controlled developmental division between long-lived highly fertile queens and short-lived functionally sterile workers (Fig. 2), it does not provide an unambiguous mechanistic explanation of this process based on defined enzymes with known properties. Given the emerging view that all DNMTs are not only functionally interweaved but also cooperate with histone modifiers, the impact of DNMT3 silencing on honeybee phenotypes is more likely the outcome of a global effect that creates a disturbance in a highly interconnected epigenomic regulatory system. For the purpose of moving forward insect epigenetics, one needs to focus on unravelling the functions of both DNA and histone modifiers before a model of this intriguing epigenetic phenomenon can be generated. In this context, the queen bee food, royal jelly (RJ), may well hold clues to this problem. RJ is a complex mixture of unique compounds with poorly understood characteristics (Maleszka 2014). Some of these compounds have been identified as histone deacetylase inhibitors and it is reasonable to assume that many other ingredients in RJ also have potent epigenetic properties, possibly affecting DNA/RNA modifying enzymology. Cloning and in vitro characterisation of the honeybee DNMTs and other genes combined with examining the effects of RJ components will provide unprecedented clues to the nature of dietary impacts on epigenetic machinery (Fig. 2).

Caste determination in *A. mellifera* is an illustrative example of Waddington's developmental canalisation, whereby epigenetic processes have been recruited to conditionally modulate the expression of one genome using an environmental cue. Although various cellular elements have the ability to respond to environmental changes, their combined and coordinated action has evolved in honeybees as a controlling mechanism for reprogramming the entire larval development with critical consequences for cellular and organismal phenotypes. A limited number of genes can be epigenetically programmed to yield more than one organismal outcome, suggesting that epigenomic modifiers are able to relax evolutionary constrains on development. These modifiers operate by recruiting only a subset of an organism's gene repertoire and reusing it in a combinatorial manner to remodel multiple subnetworks (Maleszka 2014).

Importantly, DNA methylation patterns are governed not only by the de novo and maintenance activity of DNMTs but also by the passive loss of methylation during replication and active demethylation by the family of ten eleven translocation (TET1-3) dioxygenases (Pastor et al. 2013). Demethylation by TETs is critical to ensure the flexibility of methyl marks whose responsive nature is contingent on their reversibility. These enzymes oxidise 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) that can be further converted to 5-formyl-cytosine and 5-carboxyl-cytosine. Recent findings including genome-wide mapping of 5hmC at a single-base resolution in mammalian brain (Lister et al. 2013) reveal a complex picture consistent with the idea that both 5mC and 5hmC can act as independent epigenetic marks. In addition to their role as methylcytosine dioxygenases,

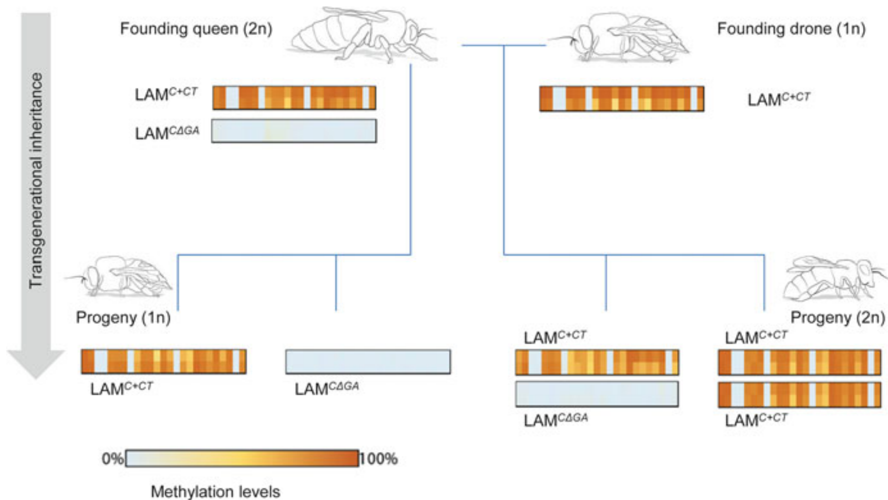


Fig. 3 Transgenerational inheritance of obligatory methylated epialleles of the gene coding for lysosomal alpha-mannosidase (LAM) in *A. mellifera*. A single male drone was mated with a queen; haploid drones develop from unfertilised eggs and diploid workers/queens develop from fertilised eggs. The level of methylation of the two alleles, LAM*C+CT* and LAM*CΔGA*, across an intragenic region of LAM (exons 16 to exons 18) is indicated: 0 % indicating a CpG site was never found to methylated, 100 % indicating that a CpG site was always found to be methylated. The transgenerational inheritance of methylation patterns, irrespective of factors such as caste or developmental stage and without parent-of-origin effects, is indicative of an obligatory methylated epiallele (see Wedd et al. 2016 for more details)

mammalian TETs perform other functions including interactions with metabolic enzymes and other proteins, participation in transcriptional regulation, telomere elongation and conveying cellular signals (Pastor et al. 2013; Lu et al. 2014). TETs have also been implicated in histone modifications. For example, TET2 promotes histone O-GlcNAcylation during TET2-dependent gene transcription by recruiting O-GlcNAc transferase to chromatin (Chen et al. 2013).

Yet, whilst active demethylation has been relatively well investigated in mammals, little is known of these processes in invertebrates. Recent work in *A. mellifera* has highlighted that a conserved TET enzyme in this organism is capable of converting methylated cytosine residues to 5-hydroxymethylcytosines (5hmC), as in mammals and is likely to play an important role in establishing DNA methylation patterns, although the exact nature of this role remains unclear (Wojciechowski et al. 2014). In contrast to mammals that have three TET paralogs, only one TET has been found in *A. mellifera* and in most of other invertebrates. TET proteins are absent in organisms such as nematodes, which have lost the genes encoding DNMT1 and DNMT3 as well as, a tRNA-methylating enzyme (known as DNMT2). One puzzling feature of the honeybee TET is its very high level of expression, especially in the brain that does not translate into a substantial number of 5hmCs. In most tissues fewer than 3000 5mhCs have been detected corresponding to 4–5 % of the

methylated cytosines. Only in testes and ovaries the levels of 5hmC in *A. mellifera* (7–10 %) appear to be more comparable to those in some mammalian situations. Given this discrepancy between TET expression and the scarcity of 5hmC in the honeybee genome, it is likely that this protein performs other roles not related to cytosine demethylation. A broader functionality of invertebrate TETs is consistent with its expression in *Drosophila* that has no DNA methylation toolkit and does not require demethylation of genomic cytosines. It is noteworthy that a few studies have reported the presence of low levels of highly localised and asymmetrical methylated cytosines in *Drosophila* genome (Takayama et al. 2014; Capuano et al. 2014). If confirmed, these results would indicate novel enzymatic machinery capable of modifying DNA in certain contexts.

A recent breakthrough study in *Drosophila* has shown that removal of the TET gene increases N6-methyladenine (m6A) levels in DNA, but has no effects on RNA in which m6A is common (Zhang et al. 2015). *Drosophila* TET null mutants show a lethal phenotype beginning at the pupal stage with the last survivors dying 3 days post-eclosion. This finding suggests that a single relative of TETs in insects has the potential to catalyse a variety of chemical modifications not necessarily involving 5mC. Even more intriguing is a new study showing that in *Drosophila*, TET has the capacity to add hydroxymethylcytosine to RNA and that this modification is most prominent in the brain where it is important for promoting translation of mRNAs (Delatte et al. 2015). Whether or not a similar TET activity exists in other insects remains to be determined, but it is conceivable that high levels of TET transcripts in *A. mellifera* brain are needed for RNA hydroxymethylation. Further exploration of demethylation and hydroxymethylation dynamics is required to fully understand both the role of TET in DNA/RNA methylation patterning and in other hitherto unidentified functions amongst invertebrate organisms.

How are DNA methyltransferases and TET dioxygenases guided to a given genomic region? Recent models indicate that these enzymes are recruited to specific genomic locations via their interactions with other epigenetic modifiers and that these interactions are dependent on the chromatin environment (Jones and Liang 2009; Jeltsch and Jurkowska 2014). Work in mammals, for instance, has highlighted that various histone post-translational modifications (PTMs) and chromatin remodelling factors are likely to be involved in recruiting DNMTs to a genomic location (Jones and Liang 2009). It is likely that similar mechanisms operate in insects, but until these are explored in the context of differential DNA methylation, the role these factors play in establishing DNA methylation patterns will remain unknown. To this end extensive PTMs have been described in *A. mellifera* using mass spectrometry (Dickman et al. 2013), and genome-wide chromatin immunoprecipitation-based approach (ChiPseq) has been applied to map-specific histone PTMs in the ant *C. floridanus* (Simola et al. 2013). Although the exact recruitment mechanism of the DNA methylation and demethylation machinery in insects is still unknown, the distribution of methylated cytosines across the genome of *A. mellifera* has been mapped at single-base resolution providing initial impetus for studies on how DNA methylation might be contributing to a variety of cellular processes.

4.2 DNA Methylation Patterns Across Invertebrates

The targeting of methylation across the genome appears to take on two distinct types of patterns: global DNA methylation and mosaic DNA methylation (Suzuki and Bird 2008). In the case of global DNA methylation, which is typical of vertebrates, the genome is densely methylated throughout most tissues and developmental stages. In humans, for instance, the majority of CpG sites across the genome are methylated with the exception of small unmethylated domains, typically near regulatory regions such as promoters, termed CpG islands (Ehrlich et al. 1982). In contrast, in invertebrate genomes, distinct domains are methylated and unmethylated in a 'mosaic' fashion (Suzuki and Bird 2008; Tweedie et al. 1997). The overall level of this mosaic methylation varies; for instance, *A. mellifera* contains a much lower level of methylation when compared to invertebrates such as *Ciona intestinalis* (Feng et al. 2010). However, several features of DNA methylation patterning are conserved.

This important epigenetic mark has long been associated with the transcriptional silencing of genomic regions. Across a wide range of organisms, it is frequently targeted to repetitive elements and transposons; *A. thaliana*, together with the mouse and zebrafish all exhibit such methylation, and disruption of DNA methylation leads to the reactivation of these elements, showing that the methylation of these regions is critical to their silencing (Bourc'his and Bestor 2004; Kato et al. 2003; Feng et al. 2010). Additionally, genomic imprinting, whereby a gene or chromosomal region are transcriptionally controlled in a parent-of-origin manner, is frequently associated with methylation and silencing. In mammals and plants imprinted genes are differentially methylated and disruption of DNA methyltransferase activity leads to the aberrant expression of the maternal and/or paternal transcript (Li et al. 1993; Jullien et al. 2006).

Yet, whilst it is clear that DNA methylation has been utilised by many organisms to maintain genome stability and regulate imprinted regions, this does not extend to all species. Whereas moderate levels of methylation have been observed across transposons in *C. intestinalis*, other invertebrate genomes do not exhibit such methylation; in *A. mellifera*, for example, transposons are not targeted for methylation (Feng et al. 2010; Zemach et al. 2010). However, transposons are not as frequent in *A. mellifera* as they are in some other species, suggesting that a mechanism controlling mobile genomic elements may be less critical for this organism. Additionally, not all species utilise DNA methylation to establish imprints; organisms such as *C. elegans* do not have a functional methylation system, yet still present with imprinted genes, and there is no current evidence for genomic imprinting via DNA methylation in *A. mellifera* (Sha and Fire 2005). It thus appears that whilst DNA methylation can be critical for transcriptional silencing in many contexts, this is not always the case.

Rather, the more conserved feature of eukaryotic DNA methylation is the distribution of methylation at gene bodies. Gene body methylation has been identified across numerous organisms, and it is likely that its occurrence predates the last common ancestor of plants and animals (Feng et al. 2010). It is characterised by

methylation across the introns and exons of protein-coding genes. In vertebrates and plants, this gene body methylation is prevalent and has been shown to frequently occur in those regions that are highly expressed (Ball et al. 2009; Zilberman et al. 2007).

In invertebrates, methylation predominantly occurs across gene bodies. The vase tunicate *C. intestinalis*, the parasitoid wasp *Nasonia vitripennis*, the carpenter ant *Camponotus floridanus*, *A.mellifera*, and lower *Metazoa* such as the sea anemone *Nematostella vectensis* all display high levels of gene body methylation, and the regions with high gene body methylation also correlate with genomic regions that are actively transcribed (Feng et al. 2010; Wang et al. 2013; Zemach et al. 2010; Foret et al. 2009; Bonasio et al. 2012; Tweedie et al. 1997). To date, in all insects in which methylation has been analysed at a genome-wide scale, 5mC appears to be limited almost exclusively to CpG dinucleotides with only marginal levels found at non-CpG sites. Whether this pattern is a universal feature of all insects awaits sequencing of methylomes in more species representing diverse evolutionary lineages. In *A. mellifera*, intragenic DNA methylation is higher in exons than introns, and those genes, which are highly methylated, are often ubiquitously expressed, with evidence that this type of methylation modulates the expression of these transcripts (Foret et al. 2009; Lyko et al. 2010). These common findings suggesting that intragenic DNA methylation is associated with active transcription, have important implications for understanding how this mark might direct transcription.

The correlation between intragenic DNA methylation and the expression of 'housekeeping' genes has led to the suggestion that DNA methylation functions in these contexts by preventing spurious transcription. By methylating intragenic regions, the initiation of transcription at cryptic sites within coding regions can be prevented (Weber and Schubeler 2007). For those genes which are frequently transcribed, such as the ubiquitously expressed housekeeping genes, this would represent an important mechanism, whereby transcription can be controlled and transcriptional noise reduced (Zilberman et al. 2007; Suzuki et al. 2007). This hypothesis is supported by a recent work, which has demonstrated that gene body methylation is negatively correlated with transcriptional noise (Huh et al. 2013). However, alternative suggestions, based on work in *A. thaliana*, indicate that it is the rate of transcription of these moderately expressed transcripts that induces intragenic methylation, as opposed to methylation itself influencing transcription (Zilberman et al. 2007).

Yet, given that key differences in the DNA methylation system exist between *A. thaliana* and other eukaryotic model organisms, the relationship between intragenic methylation and the rate of transcription may not be universal. DNA methylation patterns in *A. thaliana* and other flowering plants are established via an RNA-directed DNA methylation system (RdDM) where RNA molecules, small interfering RNAs (siRNAs), guide de novo methylation (Chan et al. 2005; Bao et al. 2004). The proposed model that transcription itself causes intragenic methylation is based upon the spurious transcription of these siRNAs (Zilberman et al. 2007). Amongst invertebrate models, there is no current evidence for an RdDM system and therefore the intragenic DNA methylation seen in these species may have alternate functions.

Table. 2 Examples of genes with cassette exons whose expression correlates with differential methylation

Gene Id	Exon	Effect on protein	Putative function
GB43824 Trans-membrane lysoplasmalogenase (Lyko et al. 2010)	3	Creates a truncated protein by introducing a premature STOP codon	In a heterodimer, the truncated protein is predicted to inhibit trans-membrane activity of the full-length protein
GB43446 Anaplastic lymphoma kinase (ALK) (Foret et al. 2012)	25	Creates a different interacting domain by changing the N-terminal sequence of ALK	Modifies selection of protein partners interacting with ALK
GB43543 Glycine receptor (Foret et al. 2012)	8	Extends the intracellular TM3-4 loop by 36 bp (12aa)	New gating properties of the glycine channel with the insertion. The cassette exon is found mostly in GlyR expressed in the brain and sensory organs

Indeed, several lines of evidence indicate a number of functional roles for intragenic DNA methylation in mammals, including the regulation of non-coding RNAs and transcription elongation. Both microRNAs (miRNAs) and long non-coding RNA (lncRNA) transcripts have been shown to be influenced by intragenic DNA methylation, with important transcriptional consequences (Kulis et al. 2013; Cheung et al. 2011; Lujambio et al. 2007). By altering chromatin structure, DNA methylation can reduce the efficiency of transcription elongation (Lorincz et al. 2004). Intragenic DNA methylation has been shown to influence elongation; exonic differential methylation modulates binding of the CTCF transcription factor, altering RNA Pol II processing and the alternative splicing of a transcript (Shukla et al. 2011). Further evidence indicates that intragenic DNA methylation modulates alternative splicing (Maunakea et al. 2013; Foret et al. 2012) and leads to the cell- and tissue-specific expression of alternative transcripts (Maunakea et al. 2010) across a number of species suggesting that this function of intragenic DNA methylation is commonplace and likely to have important biological consequences (Kulis et al. 2013).

In *A. mellifera* differential intragenic DNA methylation has similarly been correlated to alternative splicing. High intragenic DNA methylation has been observed near alternative splice sites, and the differential methylation of these regions has been linked to the expression of condition-specific alternatively spliced transcript variants (Table 2) (Lyko and Maleszka 2011; Foret et al. 2012; Kucharski et al. 2016; Maleszka et al. 2016). Some of these cases support the Shukla *et al.* (Shukla et al. 2011) model whereby methylation inhibits binding of a CTCF factor that can promote inclusion of weak upstream exons by mediating local RNA polymerase II pausing. In the case of anaplastic lymphoma kinase (Table 2), low methylation correlates with exon 25 inclusion at high frequency. In the adult brain, where the methylated regions show much lower methylation relative to larvae, the majority of the available ALK transcripts (81 %) have been found to contain exon 25 (Foret et al. 2012). Thus, differential methylation of this sequence seems to affect the dynamics of exon 25 inclusion and generates ALK isoforms with different amino-termini in

the intracellular domain that could interact with distinct partners. This suggests that the DNA methylation seen across the genes of *A. mellifera* can, in particular contexts, elicit transcriptional changes.

4.3 Does DNA Methylation Direct Gene Expression?

Although several clear relationships between DNA methylation and transcription have emerged over the past decades, including the link between intragenic DNA methylation and active transcription, this relationship is not entirely straightforward. Whilst there are a number of cases where DNA methylation has been shown to directly influence transcription, for instance, Shukla et al. (Shukla et al. 2011) provide a direct relationship between methylation and alternative splicing, much evidence is largely correlative, derived from genome-wide analyses (Zemach et al. 2010; Feng et al. 2010; Schultz et al. 2015). In the case of invertebrate models, the few examples seen in *A. mellifera* indicate that intragenic DNA methylation is likely to play an important role in eliciting transcriptional changes. Yet, these studies also highlight that such changes are context-dependent and do not necessarily eliminate the possibility that they result as a function of other processes that are simply associated with differential DNA methylation.

For some time it has been acknowledged that DNA methylation does not stand alone in directing transcriptional changes but that it is coupled with other epigenetic marks, such as histone modifications (Fuks 2005; Cedar and Bergman 2009). The epigenetic modifications and DNA methylation itself are also influenced by genetic variation (Kilpinen et al. 2013; McVicker et al. 2013; Kasowski et al. 2013; Furey and Sethupathy 2013). The complexity of these interactions makes it difficult to ascertain whether DNA methylation itself directs transcriptional changes or if other factors, such as the underlying DNA sequence, play a more substantive role, and this has become an important avenue of research across the epigenetics field (Schubeler 2012).

Cis- and *trans*-acting polymorphisms have been shown to lead to differentially methylated regions across a number of organisms (Richards 2006; Schubeler 2012). In mammals, obligatory methylated epialleles have been well documented and numerous studies have established the linkage between allelic variation and a methylated state across many cell types (Kerkel et al. 2008; Schilling et al. 2009; Shoemaker et al. 2010). Importantly, epiallelic variation is a key contributor to phenotypic variation, where these epialleles influence transcription across different cellular contexts (Gutierrez-Arcelus et al. 2013; Schubeler 2012). In *A. mellifera* and other invertebrate models, although DNA methylation is critical, there has been little investigation into how genetic variability might be contributing to differential DNA methylation.

In a recent study differential intragenic DNA methylation of the gene encoding for lysosomal alpha-mannosidase (LAM) was correlated with sequence variation, providing the first evidence for an obligatory methylated epiallele within the *A. mellifera* population (Wedd et al. 2016). In contrast to pure epialleles that arise from stochastic

genotype-independent events, obligatory differentially methylated epialleles are generated by sequence variants (Richards 2006). The methylated status of the LAM epiallele in *A. mellifera* has been found to increase LAM expression in a context-specific manner, along with the expression of a long non-coding RNA transcript. LAM epialleles have been found to be inherited in the classical Mendelian manner with no apparent evidence of imprinting (Fig. 3). This result not only further supports the link between intragenic DNA methylation and active transcription but also indicates that obligatory epialleles in *A. mellifera* will likely contribute to phenotypic variation in a context-specific manner, as has been demonstrated across other species.

Given that phenotypic differences in *A. mellifera* are generated by tightly controlled epigenetic changes, any impact that genetic variation might have on this epigenetic layer of information could be profound. In the context of DNA methylation, substantial work in social insects like *A. mellifera* has shown that differential DNA methylation patterns correlate with transcriptional changes and phenotypic variation, but fewer studies interpret these changes in the context of any underlying sequence variation. The discovery of obligatory epialleles in the *A. mellifera* population highlights the importance of interpreting differential DNA methylation patterns more carefully and investigating the extent to which epiallelic variation influences phenotype in this organism. The availability of ultra-deep next-generation bisulfite sequencing technologies, as used in the LAM study, will facilitate interpreting these patterns accurately in the context of developmental, tissue-specific and stochastic effects that are known to influence methylation patterns (Wagner et al. 2014; Landan et al. 2012).

Conclusion

Invertebrates such as *A. mellifera* represent an important model from which a broader understanding of DNA methylation, and its role in directing transcription, can be drawn (Lyko and Maleszka 2011; Maleszka 2014). Importantly, this is reliant on the accurate interpretation of insect methylomes, and without performing in-depth analyses of differential methylation patterns in *A. mellifera* there is a risk of misinterpreting the biological significance of such marks (Kucharski et al. 2015). To achieve a full understanding of how this epigenetic modification links genotype to phenotype, the relationship between DNA methylation patterns, genetic variability and other epigenetic marks must be understood and interpreted in the context of additional factors, such as developmental and tissue-specific effects and environmental influences.

References

Ball MP, Li JB, Gao Y, Lee JH, LeProust EM, Park IH, et al. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells (vol 27, pg 361, 2009). Nat Biotechnol. 2009;27(5):485.

Bao N, Lye KW, Barton MK. MicroRNA binding sites in Arabidopsis class IIIHD-ZIP mRNAs are required for methylation of the template chromosome. Dev Cell. 2004;7(5):653–62.

Bernstein BE, Meissner A, Lander ES. The mammalian epigenome. Cell. 2007;128(4):669–81.

Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH roadmap epigenomics mapping consortium. Nat Biotechnol. 2010;28(10):1045–8. doi:10.1038/nbt1010-1045.

Bird A. DNA methylation patterns and epigenetic memory. Genes Dev. 2002;16(1):6–21. doi:10.1101/Gad.947102.

Bonasio R, Li QY, Lian JM, Mutti NS, Jin LJ, Zhao HM, et al. Genome-wide and caste-specific DNA methylomes of the ants Camponotus floridanus and Harpegnathos saltator. Curr Biol. 2012;22(19):1755–64.

Bourc'his D, Bestor TH. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature. 2004;431(7004):96–9.

Capuano F, Mulleder M, Kok R, Blom HJ, Ralser M. Cytosine DNA methylation is found in Drosophila melanogaster but absent in Saccharomyces cerevisiae, Schizosaccharomyces pombe, and other yeast species. Anal Chem. 2014;86(8):3697–702. doi:10.1021/ac500447w. Epub 2014 Mar 25.

Cedar H, Bergman Y. Linking DNA methylation and histone modification: patterns and paradigms. Nat Rev Genet. 2009;10(5):295–304. doi:10.1038/nrg2540.

Chan SWL, Henderson IR, Jacobsen SE. Gardening the genome: DNA methylation in Arabidopsis thaliana. Nat Rev Genet. 2005;6(5):351–60.

Chen Q, Chen Y, Bian C, Fujiki R, Yu X. TET2 promotes histone O-GlcNAcylation during gene transcription. Nature. 2013;493(7433):561–4.

Cheung HH, Davis AJ, Lee TL, Pang AL, Nagrani S, Rennert OM, et al. Methylation of an intronic region regulates miR-199a in testicular tumor malignancy. Oncogene. 2011;30(31):3404–15.

Clamp M, Fry B, Kamal M, Xie XH, Cuff J, Lin MF, et al. Distinguishing protein-coding and noncoding genes in the human genome. Proc Natl Acad Sci U S A. 2007;104(49):19428–33. doi:10.1073/pnas.0709013104.

Delatte B, Jeschke J, Defrance M, Bachman M, Creppe C, Calonne E, et al. Genome-wide hydroxymethylcytosine pattern changes in response to oxidative stress. Sci Rep. 2015;5:12714. doi:10.1038/srep12714.

Dickman MJ, Kucharski R, Maleszka R, Hurd PJ. Extensive histone post-translational modification in honey bees. Insect Biochem Mol Biol. 2013;43(2):125–37. doi:10.1016/j.ibmb.2012.11.003.

Ehrlich M, Gamasosa MA, Huang LH, Midgett RM, Kuo KC, Mccune RA, et al. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells. Nucleic Acids Res. 1982;10(8):2709–21.

Feng SH, Cokus SJ, Zhang XY, Chen PY, Bostick M, Goll MG, et al. Conservation and divergence of methylation patterning in plants and animals. Proc Natl Acad Sci U S A. 2010;107(19): 8689–94.

Foret S, Kucharski R, Pellegrini M, Feng SH, Jacobsen SE, Robinson GE, et al. DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. Proc Natl Acad Sci U S A. 2012;109(13):4968–73. doi:10.1073/pnas.1202392109.

Foret S, Kucharski R, Pittelkow Y, Lockett GA, Maleszka R. Epigenetic regulation of the honey bee transcriptome: unravelling the nature of methylated genes. BMC Genomics. 2009;10:472.

Fuks F. DNA methylation and histone modifications: teaming up to silence genes. Curr Opin Genet Dev. 2005;15(5):490–5.

Furey TS, Sethupathy P. Genetics driving epigenetics. Science. 2013;342(6159):705–6. doi:10.1126/science.1246755.

Goll MG, Bestor TH. Eukaryotic cytosine methyltransferases. Annu Rev Biochem. 2005;74:481–514. doi:10.1146/annurev.biochem.74.010904.153721.

Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, Buil A, Ongen H, Yurovsky A, et al. Passive and active DNA methylation and the interplay with genetic variation in gene regulation, Elife. 2013;2:e00523.

Haig D. The (dual) origin of epigenetics. Cold Spring Harb Symp Quant Biol. 2004;69:67–70.

Haig D. Commentary: the epidemiology of epigenetics. Int J Epidemiol. 2012;41(1):13–6.

Halfmann R, Lindquist S. Epigenetics in the extreme: prions and the inheritance of environmentally acquired traits. Science. 2010;330(6004):629–32.

Holliday R, Pugh JE. DNA modification mechanisms for control of gene activity during development. Heredity. 1975;35(Aug):149.

Huh I, Zeng J, Park T, Yi SV. DNA methylation and transcriptional noise. Epigenetics Chromatin. 2013;6:9.

Jablonka E, Lamb MJ. The changing concept of epigenetics. From epigenesis to epigenetics: the genome in context. Ann NY Acad Sci. 2002;981:82–96.

Jablonka E, Lamm E. Commentary: the epigenotype-a dynamic network view of development. Int J Epidemiol. 2012;41(1):16–20.

Jeltsch A, Jurkowska RZ. New concepts in DNA methylation. Trends Biochem Sci. 2014; 39(7):310–8.

Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012;13(7):484–92.

Jones PA, Liang GN. OPINION Rethinking how DNA methylation patterns are maintained. Nat Rev Genet. 2009;10(11):805–11.

Jones PA, Takai D. The role of DNA methylation in mammalian epigenetics. Science. 2001; 293(5532):1068–70.

Jullien PE, Kinoshita T, Ohad N, Berger F. Maintenance of DNA methylation during the Arabidopsis life cycle is essential for parental imprinting. Plant Cell. 2006;18(6):1360–72.

Kasowski M, Kyriazopoulou-Panagiotopoulou S, Grubert F, Zaugg JB, Kundaje A, Liu YL, et al. Extensive variation in chromatin states across humans. Science. 2013;342(6159):750–2. doi:10.1126/science.1242510.

Kato M, Miura A, Bender J, Jacobsen SE, Kakutani T. Role of CG and non-CG methylation in immobilization of transposons in arabidopsis. Curr Biol. 2003;13(5):421–6.

Kerkel K, Spadola A, Yuan E, Kosek J, Jiang L, Hod E, et al. Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. Nat Genet. 2008;40(7):904–8. doi:10.1038/Ng.174.

Kilpinen H, Waszak SM, Gschwind AR, Raghav SK, Witwicki RM, Orioli A, et al. Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. Science. 2013;342(6159):744–7. doi:10.1126/science.1242463.

Kouzarides T. Chromatin modifications and their function. Cell. 2007;128(4):693–705.

Kucharski R, Foret S, Maleszka R. EGFR gene methylation is not involved in Royalactin controlled phenotypic polymorphism in honey bees. Sci Rep-Uk. 2015;5:14070. doi:10.1038/srep14070.

Kucharski R, Maleszka J, Foret S, Maleszka R. Nutritional control of reproductive status in honeybees via DNA methylation. Science. 2008;319(5871):1827–30. doi:10.1126/science.1153069.

Kucharski R, Maleszka J, Maleszka R. A possible role of DNA methylation in functional divergence of a fast evolving duplicate gene encoding odorant binding protein 11 in the honeybee. Proc Biol Sci. 2016;283(1833):20160558. doi: 10.1098/rspb.2016.0558.

Kulis M, Queiros AC, Beekman R, Martin-Subero JI. Intragenic DNA methylation in transcriptional regulation, normal differentiation and cancer. Bba-Gene Regul Mech. 2013;1829(11):1161–74.

Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518(7539):317–30.

Landan G, Cohen NM, Mukamel Z, Bar A, Molchadsky A, Brosh R, et al. Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. Nat Genet. 2012;44(11):1207–14. doi:10.1038/ng.2442.

Li E, Beard C, Jaenisch R. Role for DNA methylation in genomic imprinting. Nature. 1993;366(6453):362–5.

Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, et al. Global epigenomic reconfiguration during mammalian brain development. Science. 2013;341(6146):629–40. doi:10.1126/science.1237905.

Lorincz MC, Dickerson DR, Schmitt M, Groudine M. Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. Nat Struct Mol Biol. 2004;11(11):1068–75. doi:10.1038/Nsmb840.

Lu FL, Liu YT, Jiang L, Yamaguchi S, Zhang Y. Role of Tet proteins in enhancer activity and telomere elongation. Genes Dev. 2014;28(19):2103–19.
Lujambio A, Ropero S, Ballestar E, Fraga MF, Cerrato C, Setien F, et al. Genetic unmasking of an epigenetically silenced microRNA in human cancer cells. Cancer Res. 2007;67(4):1424–9.
Lyko F, Foret S, Kucharski R, Wolf S, Falckenhayn C, Maleszka R. The honey bee epigenomes: differential methylation of brain DNA in queens and workers. PLoS Biol. 2010;8(11):e1000506. doi:10.1371/journal.pbio.1000506.
Lyko F, Maleszka R. Insects as innovative models for functional studies of DNA methylation. Trends Genet. 2011;27(4):127–31. doi:10.1016/j.tig.2011.01.003.
Maleszka R. Epigenetic integration of environmental and genomic signals in honey bees: the critical interplay of nutritional, brain and reproductive networks. Epigenetics. 2008;3(4):188–92: 6697 [pii].
Maleszka R. The social honey bee in biomedical research: realities and expectations. Drug Discov Today Dis Models. 2014;12:7–13. doi:10.1016/j.ddmod.2014.06.001.
Maleszka R, Mason PH, Barron AB. Epigenomics and the concept of degeneracy in biological systems. Brief Funct Genomics. 2014;13(3):191–202. doi:10.1093/bfgp/elt050.
Maleszka R. Epigenetic code and insect behavioural plasticity. Curr Opin Insect Sci. 2016;15:45–52. doi:10.1016/j.cois.2016.03.003.
Maunakea AK, Chepelev I, Cui KR, Zhao KJ. Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. Cell Res. 2013;23(11):1256–69.
Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature. 2010; 466(7303):253–7.
McVicker G, van de Geijn B, Degner JF, Cain CE, Banovich NE, Raj A, et al. Identification of genetic variants that affect histone modifications in human cells. Science. 2013;342(6159):747–9. doi:10.1126/science.1242429.
Miklos GLG, Maleszka R. Epigenomic communication systems in humans and honey bees: from molecules to behavior. Horm Behav. 2011;59(3):399–406. doi:10.1016/j.yhbeh.2010.05.016.
Pastor WA, Aravind L, Rao A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. Nat Rev Mol Cell Biol. 2013;14(6):341–56. doi:10.1038/Nrm3589.
Regev A, Lamb MJ, Jablonka E. The role of DNA methylation in invertebrates: developmental regulation or genome defense? Mol Biol Evol. 1998;15(7):880–91.
Richards EJ. Opinion – inherited epigenetic variation – revisiting soft inheritance. Nat Rev Genet. 2006;7(5):395–401. doi:10.1038/Nrg1834.
Riggs AD. X-inactivation, differentiation, and DNA methylation. Cytogenet Cell Genet. 1975; 14(1):9–25.
Schilling E, El Chartouni C, Rehli M. Allele-specific DNA methylation in mouse strains is mainly determined by cis-acting sequences. Genome Res. 2009;19(11):2028–35. doi:10.1101/gr.095562.109.
Schubeler D. Epigenetic islands in a genetic ocean. Science. 2012;338(6108):756–7. doi:10.1126/science.1227243.
Schultz MD, He YP, Whitaker JW, Hariharan M, Mukamel EA, Leung D, et al. Human body epigenome maps reveal noncanonical DNA methylation variation. Nature. 2015;523(7559):212–6.
Sha K, Fire A. Imprinting capacity of gamete lineages in Caenorhabditis elegans. Genetics. 2005;170(4):1633–52.
Shoemaker R, Deng J, Wang W, Zhang K. Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. Genome Res. 2010;20(7):883–9. doi:10.1101/gr.104695.109.
Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature. 2011;479(7371):74–9. doi:10.1038/Nature10442.
Simola DF, Ye C, Mutti NS, Dolezal K, Bonasio R, Liebig J, et al. A chromatin link to caste identity in the carpenter ant Camponotus floridanus. Genome Res. 2013;23(3):486–96.

Smith ZD, Meissner A. DNA methylation: roles in mammalian development. Nat Rev Genet. 2013;14(3):204–20. doi:10.1038/nrg3354.
Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet. 2008;9(6):465–76.
Suzuki MM, Kerr ARW, De Sousa D, Bird A. CpG methylation is targeted to transcription units in an invertebrate genome. Genome Res. 2007;17(5):625–31.
Takayama S, Dhahbi J, Roberts A, Mao G, Heo SJ, Pachter L, et al. Genome methylation in D. melanogaster is found at specific short motifs and is independent of DNMT2 activity. Genome Res. 2014;24(5):821–30.
Tweedie S, Charlton J, Clark V, Bird A. Methylation of genomes and genes at the invertebrate-vertebrate boundary. Mol Cell Biol. 1997;17(3):1469–75.
Waddington CH. The strategy of the genes. A discussion of some aspects of theoretical biology. London: George Allen & Unwin; 1957. p. 1957.
Wagner JR, Busche S, Ge B, Kwan T, Pastinen T, Blanchette M. The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. Genome Biol. 2014;15(2):R37. doi:10.1186/gb-2014-15-2-r37.
Wang X, Wheeler D, Avery A, Rago A, Choi JH, Colbourne JK, et al. Function and evolution of DNA methylation in Nasonia vitripennis. PLoS Genet. 2013;9(10):e1003872.
Weber M, Schubeler D. Genomic patterns of DNA methylation: targets and function of an epigenetic mark. Curr Opin Cell Biol. 2007;19(3):273–80.
Wedd L, Kucharski R, Maleszka R. Differentially methylated obligatory epialleles modulate context-dependent LAM gene expression in the honey bee Apis mellifera. Epigenetics. 2016;11(1):1–10.
Wojciechowski M, Rafalski D, Kucharski R, Misztal K, Maleszka J, Bochtler M, et al. Insights into DNA hydroxymethylation in the honeybee from in-depth analyses of TET dioxygenase. Open Biol. 2014;4(8). doi:Unsp 140110.
Xiang H, Zhu JD, Chen QA, Dai FY, Li X, Li MW, et al. Single base-resolution methylome of the silkworm reveals a sparse epigenomic map (vol 28, pg 516, 2010). Nat Biotechnol. 2010;28(7):756.
Zemach A, McDaniel IE, Silva P, Zilberman D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science. 2010;328(5980):916–9. doi:10.1126/science.1186366.
Zhang GQ, Huang H, Liu D, Cheng Y, Liu XL, Zhang WX, et al. N-6-methyladenine DNA modification in Drosophila. Cell. 2015;161(4):893–906.
Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nat Genet. 2007;39(1):61–9.

N6-Methyladenine: A Conserved and Dynamic DNA Mark

Zach Klapholz O'Brown and Eric Lieberman Greer

Abstract

Chromatin, consisting of deoxyribonucleic acid (DNA) wrapped around histone proteins, facilitates DNA compaction and allows identical DNA codes to confer many different cellular phenotypes. This biological versatility is accomplished in large part by posttranslational modifications to histones and chemical modifications to DNA. These modifications direct the cellular machinery to expand or compact specific chromatin regions and mark regions of the DNA as important for cellular functions. While each of the four bases that make up DNA can be modified (Iyer et al. 2011), this chapter will focus on methylation of the sixth position on adenines (6mA), as this modification has been poorly characterized in recently evolved eukaryotes, but shows promise as a new conserved layer of epigenetic regulation. 6mA was previously thought to be restricted to unicellular organisms, but recent work has revealed its presence in metazoa. Here, we will briefly describe the history of 6mA, examine its evolutionary conservation, and evaluate the current methods for detecting 6mA. We will discuss the enzymes that bind and regulate this mark and finally examine known and potential functions of 6mA in eukaryotes.

Z.K. O'Brown • E.L. Greer (✉)
Division of Newborn Medicine, Boston Children's Hospital,
300 Longwood Avenue, Boston, MA 02115, USA

Department of Pediatrics, Harvard Medical School, Boston, MA 02115, USA
e-mail: eric.greer@childrens.harvard.edu

© Springer International Publishing Switzerland 2016

A. Jeltsch, R.Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*,
Advances in Experimental Medicine and Biology 945,
DOI 10.1007/978-3-319-43624-1_10

1 Introduction

DNA must faithfully transmit the blueprints of life from generation to generation. However, it is also necessary that different cell types have access to different portions of the genome and that specific cell types can respond appropriately to changes in the environment. Such dynamic responses are mediated in part by transcription factor complexes and by chemical modifications to chromatin. DNA is not as heavily modified as RNA, which has 141 different modifications identified to date (Machnicka et al. 2013; Grosjean 2015). The limited number of DNA modifications (relative to RNA) is presumably evolutionarily selected for to protect the DNA code from mutations and to enable formation of the double helix. Nevertheless, several DNA modifications occur across the tree of life and are important as both signals of DNA lesions and as epigenetic regulators of diverse biological processes. Importantly, DNA modifications increase the repertoire of cellular phenotypes that can be encoded by a single DNA sequence, without directly altering the integrity of the genetic code. Soon after DNA was discovered, variants of each base were identified. However, the role of DNA methylation in the context of normal biological processes and disease pathogenesis remains an active area of research.

Although 6mA was discovered soon after cytosine methylation (5mC), it was thought to exist predominantly in prokaryotes and was therefore not given the same amount of research attention in eukaryotes as 5mC. The discovery that 6mA exists in more recently evolved eukaryotes has revived interest in this DNA modification. To understand the dynamic regulation of and by adenine methylation, it is useful to view the role of 6mA across evolution. Here, we aim to provide a broad overview of the historical research on 6mA across the evolutionary spectrum and discuss the mechanisms by which N6-adenine methylation is established, removed, and recognized. We examine 6mA's role in biology, discuss the possibility of 6mA maintaining epigenetic information across cell divisions and potentially across generations, and summarize exciting areas for future research.

2 Types of DNA Modifications

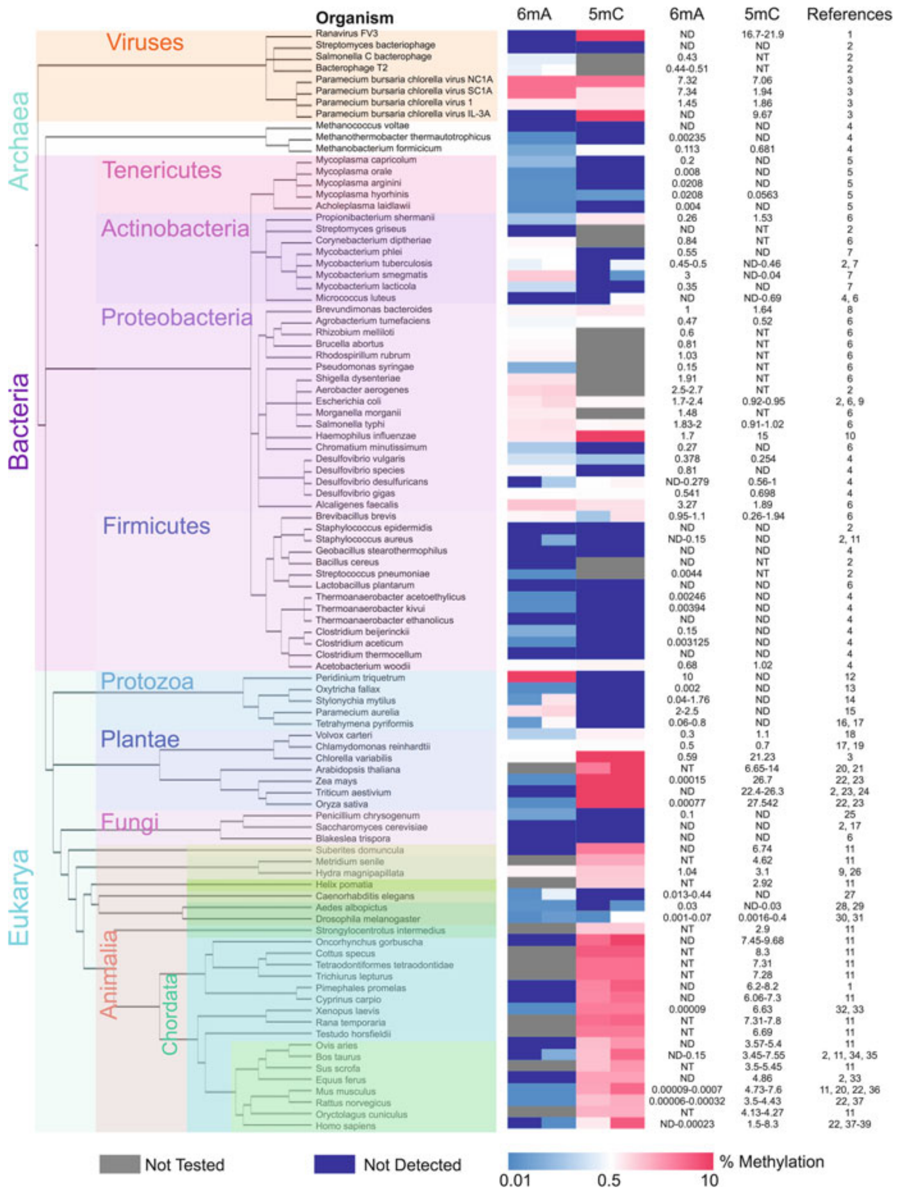
Each DNA base is modified to varying degrees in different organisms. DNA methylation occurs either as nonenzymatic DNA damaging lesions or as directed modifications with signaling function, which are actively introduced by specific methyltransferase enzymes. DNA lesions include N1-methyladenine (1mA), N3-methyladenine (3mA), N7-methyladenine (7mA), N3-methylcytosine (3mC), N2-methylguanine (2mG), O6-methylguanine (6mG), N7-methylguanine (7mG), N3-methylthymine (3mT), and O4-methylthymine (4mT), while directed methylation includes N6-methyladenine (6mA), N4-methylcytosine (4mC), and C5-methylcytosine (5mC) (Sedgwick et al. 2007; Iyer et al. 2011; Grosjean 2009). Other DNA modifications include deaminated cytosines (Shapiro and Klein 1966; Lindahl and Nyberg 1974), oxidized derivatives of 5mC (5hmC, 5fC, and 5caC) (Wyatt and Cohen 1952; Privat and Sowers 1996; Shen et al. 2014), and the hypermodified thymine base J

(Gommers-Ampt et al. 1993). These modifications are discussed in greater detail in other reviews; we will focus on 6mA, a relatively uncharacterized DNA modification in eukaryotes with potential epigenetic function.

Of the directed DNA methylation events, 5mC is the most extensively studied. 5mC occurs at a higher frequency in more recently evolved organisms and its abundance in the genome ranges from 0.002 to 27 % of cytosines, depending on the organism (Fig. 1). In mammals and plants, 5mC is the most abundant DNA modification (Iyer et al. 2011) and functions in the regulation of gene expression and maintenance of epigenetic memory (Bird 2002). 5mC in promoter regions typically leads to transcriptional gene silencing and therefore plays important roles in diverse cellular and developmental processes, including X-chromosome inactivation, genomic imprinting, stem cell pluripotency, and differentiation (Bird 2002). Other directed DNA methylation events include 4mC and 6mA. 4mC has been identified mainly in thermophilic bacteria and archaea (Janulaitis et al. 1983; Ehrlich et al. 1985, 1987; Grosjean 2009). Until recently, 6mA was also thought to be restricted to bacteria, archaea, and protists. However, its recent identification in several eukaryotes raises the possibility that 6mA serves as an epigenetic signaling modification within these organisms and potentially across generations.

3 Discovery of 6mA Across Eukaryotes

DNA N6-methyladenine (6mA) is a widespread modification in prokaryotes. Although 6mA is not necessary for viability in prokaryotes (Marinus and Morris 1973; Russell and Hirata 1989), it plays crucial roles in regulating DNA replication (Campbell and Kleckner 1990; Yamaki et al. 1988), repair (Pukkila et al. 1983), transposition (Roberts et al. 1985), transcription (Wallecha et al. 2002; Robbins-Manke et al. 2005), and cellular defense (Luria and Human 1952; Meselson and Yuan 1968; Linn and Arber 1968; Smith et al. 1972). For reviews on 6mA in prokaryotes, please see (Marinus and Lobner-Olesen 2014; Wion and Casadesus 2006; Murray 2002). The discovery of 6mA started with the identification of an unknown base initially in *E. coli*. Using several techniques, this base was compared to synthesized nucleotides to identify 6mA. Hydrolyzed bases were separated by two-dimensional paper chromatography in different solvents, ultraviolet absorption spectrum maxima and minima were measured, and electrophoretic mobility of this unknown base all confirmed the detection of 6mA (Dunn and Smith 1955, 1958). The existence of 6mA was subsequently confirmed in a variety of different bacterial species (Vanyushin et al. 1968). These initial detection techniques were capable of detecting 6mA at ~0.01 % of total adenines (Vanyushin et al. 1970). This detection limit, combined with the confounding presence of commensal symbionts, technical variability, tissue-specific differences, development/stage-specific variability, or subtle environmental effects on 6mA levels, initially led to contradictory reports of the identification of 6mA in eukaryotes. Indeed, 6mA was reported by one group to occur in bull and human sperm (Unger and Venner 1966), but other groups were unable to replicate this result or detect 6mA in other metazoa (Dunn and Smith 1958;


Viruses
Archaea
Tenericutes
Actinobacteria
Proteobacteria
Bacteria
Firmicutes
Protozoa
Plantae
Fungi
Eukarya
Animalia
Chordata
Organism 6mA 5mC 6mA 5mC References
Ranavirus FV3 ND 16.7-21.9 1
Streptomyces bacteriophage ND ND 2
Salmonella C bacterophage 0.43 NT 2
Bacterophage T2 0.44-0.51 NT 2
Paramecium bursaria chlorella virus NC1A 7.32 7.06 3
Paramecium bursaria chlorella virus SC1A 7.34 1.94 3
Paramecium bursaria chlorella virus 1 1.45 1.86 3
Paramecium bursaria chlorella virus IL-3A ND 9.67 3
Methanococcus voltae ND ND 4
Methanothermobacter thermautotrophicus 0.00235 ND 4
Methanobacterium formicicum 0.113 0.681 4
Mycoplasma capricolum 0.2 ND 5
Mycoplasma orale 0.008 ND 5
Mycoplasma arginini 0.0208 ND 5
Mycoplasma hyorhinis 0.0208 0.0563 5
Acholeplasma laidlawii 0.004 ND 5
Propionibacterium shermanii 0.26 1.53 6
Streptomyces griseus ND NT 2
Corynebacterium diptheriae 0.84 NT 6
Mycobacterium phlei 0.55 ND 7
Mycobacterium tuberculosis 0.45-0.5 ND-0.46 2, 7
Mycobacterium smegmatis 3 ND-0.04 7
Mycobacterium lacticola 0.35 ND 7
Micrococcus luteus ND ND-0.69 4, 6
Brevundimonas bacteroides 1 1.64 8
Agrobacterium tumefaciens 0.47 0.52 6
Rhizobium meliloti 0.6 NT 6
Brucella abortus 0.81 NT 6
Rhodospirillum rubrum 1.03 NT 6
Pseudomonas syringae 0.15 NT 6
Shigella dysenteriae 1.91 NT 6
Aerobacter aerogenes 2.5-2.7 NT 2
Escherichia coli 1.7-2.4 0.92-0.95 2, 6, 9
Morganella morganii 1.48 NT 6
Salmonella typhi 1.83-2 0.91-1.02 6
Haemophilus influenzae 1.7 15 10
Chromatium minutissimum 0.27 ND 6
Desulfovibrio vulgaris 0.378 0.254 4
Desulfovibrio species 0.81 ND 4
Desulfovibrio desulfuricans ND-0.279 0.56-1 4
Desulfovibrio gigas 0.541 0.698 4
Alcaligenes faecalis 3.27 1.89 6
Brevibacillus brevis 0.95-1.1 0.26-1.94 6
Staphylococcus epidermidis ND ND 2
Staphylococcus aureus ND-0.15 ND 2, 11
Geobacillus stearothermophilus ND ND 4
Bacillus cereus ND NT 2
Streptococcus pneumoniae 0.0044 NT 2
Lactobacillus plantarum ND ND 6
Thermoanaerobacter acetoethylicus 0.00246 ND 4
Thermoanaerobacter kivui 0.00394 ND 4
Thermoanaerobacter ethanolicus ND ND 4
Clostridium beijerinckii 0.15 ND 4
Clostridium aceticum 0.003125 ND 4
Clostridium thermocellum ND ND 4
Acetobacterium woodii 0.68 1.02 4
Peridinium triquetrum 10 ND 12
Oxytricha fallax 0.002 ND 13
Stylonychia mytilus 0.04-1.76 ND 14
Paramecium aurelia 2-2.5 ND 15
Tetrahymena pyriformis 0.06-0.8 ND 16, 17
Volvox carteri 0.3 1.1 18
Chlamydomonas reinhardtii 0.5 0.7 17, 19
Chlorella variabilis 0.59 21.23 3
Arabidopsis thaliana NT 6.65-14 20, 21
Zea mays 0.00015 26.7 22, 23
Triticum aestivium ND 22.4-26.3 2, 23, 24
Oryza sativa 0.00077 27.542 22, 23
Penicillium chrysogenum 0.1 ND 25
Saccharomyces cerevisiae ND ND 2, 17
Blakeslea trispora ND ND 6
Suberites domuncula ND 6.74 11
Metridium senile NT 4.62 11
Hydra magnipapillata 1.04 3.1 9, 26
Helix pomatia NT 2.92 11
Caenorhabditis elegans 0.013-0.44 ND 27
Aedes albopictus 0.03 ND-0.03 28, 29
Drosophila melanogaster 0.001-0.07 0.0016-0.4 30, 31
Strongylocentrotus intermedius NT 2.9 11
Oncorhynchus gorbuscha ND 7.45-9.68 11
Cottus specus NT 8.3 11
Tetraodontiformes tetraodontidae NT 7.31 11
Trichiurus lepturus NT 7.28 11
Pimephales promelas ND 6.2-8.2 1
Cyprinus carpio ND 6.06-7.3 11
Xenopus laevis 0.00009 6.63 32, 33
Rana temporaria NT 7.31-7.8 11
Testudo horsfieldii NT 6.69 11
Ovis aries ND 3.57-5.4 11
Bos taurus ND-0.15 3.45-7.55 2, 11, 34, 35
Sus scrofa NT 3.5-5.45 11
Equus ferus ND 4.86 2, 33
Mus musculus 0.00009-0.0007 4.73-7.6 11, 20, 22, 36
Rattus norvegicus 0.00006-0.00032 3.5-4.43 22, 37
Oryctolagus cuniculus NT 4.13-4.27 11
Homo sapiens ND-0.00023 1.5-8.3 22, 37-39
Not Tested
Not Detected
% Methylation
0.01
0.5
10

Vanyushin et al. 1970). 6mA was reported to occur in some unicellular eukaryotes including *Paramecium aurelia* (Cummings et al. 1974), *Stylonychia mytilus* (Ammermann et al. 1981), *Oxytricha fallax* (Rae and Spear 1978), *Chlorella variabilis* (Van Etten et al. 1985), *Tetrahymena pyriformis* (Gorovsky et al. 1973), and *Chlamydomonas reinhardtii* (Hattman et al. 1978). Two reports also identified 6mA in multicellular eukaryotes, including the mosquito *Aedes albopictus* (Adams et al. 1979) and the sponge *Suberites domuncula* (Vanyushin et al. 1970). However, detection of 6mA in mosquitoes was not reproduced (Proffitt et al. 1984), and its detection in the sponge was dismissed as potentially coming from symbiotic prokaryotes or algae (Vanyushin et al. 1970). Therefore, until recently, 6mA was thought to be restricted to prokaryotes and unicellular eukaryotes (Casadesus and Low 2006).

With the advent of more sensitive detection techniques (discussed below), 6mA has recently been identified in multicellular eukaryotes including *Caenorhabditis elegans* and *Drosophila melanogaster* (Greer et al. 2015b; Zhang et al. 2015). Several other papers reported low levels of 6mA in more recently evolved eukaryotes, but each of these has caveats that we must acknowledge. 6mA was detected in *Drosophila*, calf thymus, and human placental samples by dot blots (Achwal et al. 1983). A recent paper detected 6mA by immunofluorescence in mouse heart tissues (Sun et al. 2015). Another group identified 6mA in the plants *Oryza sativa* and *Zea mays*, rat tissues, and

Fig. 1 Abundance of 6mA and 5mC across the tree of life. The relative abundance of 6mA and 5mC is displayed in a heat map. The first column of the heat map displays the percentage of adenines that are N6-methylated (%6mA/A) and the second column displays the percentage of cytosines that are C5-methylated (%5mC/C) for the organism indicated in each row. *Blue color* represents lower 6mA or 5mC abundance and red color represents higher 6mA or 5mC abundance. *Gray color* indicates that the methylation mark was not tested in that organism. *Dark blue* color indicates that the methylation mark was not detected in that organism and therefore may or may not be present at levels below the limit of detection for the technique used. For some organisms, the level of methylation has been shown to vary across multiple measurements, between different studies or between different cell types within the same organism. In such cases, a range is presented where the left half of the column reflects the lowest detected level (or not detected in some cases) and the right half of the column shows the highest detected level. Methylation values are presented on the right along with citations. The phylogenetic tree was generated using the PhyloT web server (http://phylot.biobyte.de/index.html) and visualized using the interactive tree of life web server (http://itol.embl.de/). The phylogenetic tree ("rooted" setting) displays the inferred evolutionary relationships between the indicated genera based on their genetic similarity (Letunic and Bork 2011). The tree was created using FigTree v1.4.2. The different organisms are subdivided into different colored boxes to represent different kingdoms and phyla. For some phyla only one organism has been examined. 1 (Willis and Granoff 1980), 2 (Dunn and Smith 1958), 3 (Van Etten et al. 1985), 4 (Ehrlich et al. 1985), 5 (Razin and Razin 1980), 6 (Vanyushin et al. 1968), 7 (Srivastava et al. 1981), 8 (Degnen and Morris 1973), 9 (Yuki et al. 1979), 10 (Drozdz et al. 2012), 11 (Vanyushin et al. 1970), 12 (Rae 1976), 13 (Rae and Spear 1978), 14 (Ammermann et al. 1981), 15 (Cummings et al. 1974), 16 (Gorovsky et al. 1973), 17 (Hattman et al. 1978), 18 (Babinger et al. 2001), 19 (Fu et al. 2015), 20 (Capuano et al. 2014), 21 (Kakutani et al. 1999), 22 (Huang et al. 2015), 23 (Wagner and Capesius 1981), 24 (Montero et al. 1992), 25 (Rogers et al. 1986), 26 (Hassel et al. 2010), 27 (Greer et al. 2015b), 28 (Adams et al. 1979), 29 (Proffitt et al. 1984), 30 (Zhang et al. 2015), 31 (Lyko et al. 2000), 32 (Koziol et al. 2016), 33 (Jabbari et al. 1997), 34 (Unger and Venner 1966), 35 (Romanov and Vanyushin 1981), 36 (Wu et al. 2016), 37 (Gama-Sosa et al. 1983), 38 (Tawa et al. 1992), and 39 (Ehrlich et al. 1982)

human cells by high-performance liquid chromatography coupled with mass spectrometry (HPLC-ms/ms) (Huang et al. 2015). More recently, 6mA was found by dot blots, HPLC, and methyl DNA immunoprecipitation followed by sequencing (MeDIP-seq) in *Xenopus laevis* and mouse kidney (Koziol et al. 2016) and by dot blots, MeDIP-seq, HPLC, and SMRT-seq in mouse embryonic stem (ES) cells (Wu et al. 2016). While these papers raise the exciting possibility that 6mA may indeed be present across the tree of life, it is difficult to discount potential contaminating microbiota and to confirm that the detection of 6mA is real when the reported levels of 6mA are at the limit of detection. RNA m6A (discussed below) could also account for contaminating signal in dot blots and immunofluorescence if not properly removed. We must also recognize that the injection of N6-adenine methylated oligos into mice induces a greater immune response than unmethylated oligos, as measured by the production of IL-12 (Tsuchiya et al. 2005). But this does not necessarily confirm that 6mA is a foreign base in mice as unmethylated CpG motifs also induce a more substantial immune response (Tsuchiya et al. 2005). These results raise the possibility that 6mA is either not present in mammals or present in sufficiently small quantities to keep it as an immunogenic species in the mammalian repertoire. To confirm the existence of 6mA across eukaryotes, it will be necessary to identify the enzymes that regulate 6mA and biological conditions under which the modification changes. The recent studies suggest that 6mA might be a conserved DNA modification and raise several fundamental and largely unexplored questions about the evolutionary importance of 6mA across the tree of life. From an evolutionary perspective, why did higher eukaryotes shift from 6mA (the most pervasive DNA modification in prokaryotes) toward using 5mC as the more dominant DNA modification? To what extent are the ancient functions of 6mA and its modifying enzymes conserved from prokaryotes to more recent eukaryotes?

In contrast to DNA adenine methylation, RNA adenine methylation (m6A) has long been recognized as the most abundant posttranscriptional modification of prokaryotic and eukaryotic mRNAs (Niu et al. 2013). In humans, there are over 18,000 m6A sites representing approximately 7000 unique mRNA transcripts (Jia et al. 2011; Meyer et al. 2012; Dominissini et al. 2012). Furthermore, m6A is enriched in 3′UTRs in highly conserved regions (Meyer et al. 2012; Dominissini et al. 2012; Deng et al. 2015), suggesting a shared function for m6A in evolutionarily distant species. N6-methyladenosine regulates multiple aspects of RNA metabolism, including mRNA stability/decay, translation, splicing, and localization (Wang et al. 2014, 2015; Zhou et al. 2015; Niu et al. 2013), and participates in diverse cellular and biological processes including meiosis and embryonic stem cell differentiation (Yue et al. 2015; Batista et al. 2014; Hongay and Orr-Weaver 2011; Bodi et al. 2012). The prevalence of RNA m6A raises the possibility that DNA adenine methylation could be a consequence of methylated adenines in RNA recycled via the nucleotide salvage pathway. Another possibility is that DNA adenine methylation is catalyzed by RNA methyltransferases, either as an off-target effect of these enzymes or as a biologically regulated process. Unlike the better-characterized RNA m6A, relatively little is known about the functional importance of DNA 6mA in metazoan genomes and whether 6mA plays a similarly conserved role in the dynamic regulation of biological processes. The phenotypic consequences of RNA m6A might provide clues to the roles of N6-adenine methylation on DNA.

4 Abundance of 6mA

The relative genomic abundance of 6mA can provide clues to its biological relevance across evolutionarily distinct organisms. 6mA and 5mC appear to have a large range of abundance in the genomes of different organisms across evolution (Gommers-Ampt and Borst 1995). 5mC is undetectable in many bacterial species, as well as the genome of *S. cerevisiae*, and ranges from 0.0016 % of cytosines in *D. melanogaster* to as high as 10 % in some mammals and 30 % in certain plant species (Gommers-Ampt and Borst 1995; Capuano et al. 2014; Wagner and Capesius 1981). If we accept that published literature documenting the presence of 6mA in different species is in fact detecting 6mA in the reported organism (rather than in contaminating symbionts), the genomic abundance of 6mA varies by several orders of magnitude across the tree of life (Fig. 1). Generally, organisms with higher levels of 6mA such as bacteria and single-celled eukaryotes tend to have lower levels of 5mC, while organisms with higher levels of 5mC such as plants and mammals tend to have lower levels of 6mA. The detected level of 6mA ranges from ~0.0001 to 0.0003 % of adenines in plants and mammals to as high as 3 % of adenines in some species of bacteria and up to 10 % of adenines in the dinoflagellate *Peridinium triquetrum* (Rae 1976). Early studies of nucleic acid composition in the 1950s examined the base composition of DNA in different strains of bacteria using 2D paper chromatography (Dunn and Smith 1958). It was found that 6mA comprised 1.75 % of all adenines in *E. coli* and 2.5 % of adenines in *Aerobacter aerogenes* (Dunn and Smith 1958). Subsequent studies examined the content of 6mA in the DNA of unicellular eukaryotes, such as the ciliate *Tetrahymena pyriformis* (0.65–0.8 % of adenines) (Gorovsky et al. 1973), *Paramecium aurelia* (2.5 %) (Cummings et al. 1974), and *Stylonychia mytilus* (0.176 %) (Ammermann et al. 1981). The level of 6mA in these unicellular eukaryotes is comparable to the 6mA abundance in many species of bacteria. Interestingly, *Tetrahymena* and *Stylonychia mytilus* have 4–13-fold lower 6mA levels in their micronucleus than their macronucleus (Gorovsky et al. 1973; Ammermann et al. 1981), suggesting that this modification plays an important role in determining the differences between the two nuclei in these species, which are separated by ~1159 million years of evolution (Parfrey et al. 2011).

Recently, 6mA was identified in the DNA of *C. elegans*, using both antibody-based approaches and antibody-independent methods of quantitation, including single-molecule real-time (SMRT) sequencing and ultrahigh-performance liquid chromatography followed by mass spectrometry (UHPLC-ms/ms) (Greer et al. 2015b). Based on the UHPLC-ms/ms data, the levels of 6mA ranged from 0.013 to 0.39 % of adenines, representing a 30-fold variation in the global level of adenine methylation between different batches of wild-type *C. elegans*. The observation that 6mA abundance can vary by more than an order of magnitude within an isogenic population of animals is interesting, as it suggests that the levels of 6mA in these organisms might be particularly sensitive to subtle changes in the environment (e.g., stress stimuli).

A recent study quantified the genomic abundance of 6mA in plants, rat tissues, and human cells using HPLC-ms/ms (Huang et al. 2015). These data must be viewed with caution, as there was no independent validation that the 6mA

modification was occurring in the reported organisms rather than contaminating symbionts. In that study, the abundance of 6mA in plant and mammalian genomes ranged from 0.00008 % of adenines in rat lung DNA to as high as 0.0007 % of adenines in plant DNA. The human cell lines had 0.0017 % and 0.0023 % 6mA (in Jurkat and 293 T cells, respectively). Another group identified 6mA in 0.00009 % of adenines in *Xenopus laevis* by HPLC and MeDIP-seq (Koziol et al. 2016). More recently 6mA was identified in mouse ES cells at 0.0006–0.0007 % of adenines (Wu et al. 2016). These finding suggests that 6mA in plants and mammalian genomes is ~1000-40,000-fold lower in abundance than in some bacteria and single-celled eukaryotes. The large degree of variability in 6mA abundance between eukaryotes motivates further exploration into the environmental factors and evolutionary pressures that led to a decline in 6mA levels and an increase in 5mC levels during eukaryotic evolution. These differences could also indicate that at very low 6mA levels, 6mA is at the limit of detection. Therefore, quantitative differences between different samples could be attributed to technical errors rather than true biological variability. Moreover, these modifications are typically detected under basal conditions. It is possible that 6mA levels are dramatically altered under specific environmental conditions. Finally, we should note that even if a relatively rare percentage of adenines are methylated, the presence of a single methylated adenine at a critical genomic location could have dramatic phenotypic consequences by affecting the binding of specific regulatory proteins (see Sect. 9.6 below).

5 Methods of Detecting 6mA

Detection of DNA methylation has evolved over the years to become increasingly sensitive and accurate. Detecting different DNA modifications started with a technique of combining the cytosine fraction with picric acid to form crystalline picrate. After purification by crystallization, salt crystals were compared to synthetic pyrimidines of known structure. By this method, the authors reported the identification of 5mC in *Mycobacterium tuberculosis* in 1925 (Johnson and Coghill 1925). Detection techniques shifted to paper chromatography (Hotchkiss 1948), which had a limit of detection of 1 % and was used to compare synthetically generated 5mC to the content of 5mC in animal, plant, viral, and bacterial DNA (Wyatt 1950). By the time 6mA was first identified in 1955, its presence was confirmed by the analysis of ultraviolet absorption spectra (Mason 1954), electrophoretic mobility, and its paper chromatographic movement in different solvents (Dunn and Smith 1955). Because these early methods were relatively insensitive, the presence of 6mA in a number of animal species was undetectable. Researchers quickly realized that they could take advantage of restriction enzymes to identify methylated residues (Bird and Southern 1978; Geier and Modrich 1979). A limitation of this approach is that the detection of methylation sites is dependent on the methylated residue occurring in the appropriate restriction enzyme target motif and whether the restriction enzyme preferentially recognizes unmethylated, hemimethylated, or fully methylated substrates. Therefore, not all sequence contexts can be addressed with this method.

High-performance liquid chromatography was subsequently used to determine that *E. coli* has 1.4 % 6mA (Yuki et al. 1979). Liquid chromatography has become increasingly sensitive, and, recently, ultrahigh-performance liquid chromatography coupled with mass spectrometry (UHPLC-ms/ms) has been used to detect concentrations of 6mA on the order of 0.00001 % (Huang et al. 2015). An alternative technique, called capillary electrophoresis and laser-induced fluorescence (CE-LIF), uses the fluorescent dye boron-dipyrromethene (BODIPY), to specifically bind to 6mA, followed by capillary electrophoresis combined with laser-induced fluorescence to detect 6mA levels (Krais et al. 2010). This technique has a lower limit of detection of 0.01 % 6mA and was used to confirm the presence of 6mA in bacteriophage λ, *E. coli*, and to identify 6mA's presence in *Hydra magnipapillata* (1.04 % of adenines) (Krais et al. 2010). At this limit of detection, the authors could not detect 6mA in calf thymus or human kidney samples.

While the aforementioned techniques have proven useful for detecting whether 6mA is present in a particular organism, they do not provide information on the genomic location of this modification. To determine the genomic locations of 6mA, several methylation-sensitive sequencing techniques have been developed. Methylated DNA immunoprecipitation (MeDIP) coupled with microarray analysis (Weber et al. 2005) has evolved into MeDIP sequencing (MeDIP-seq) (Pomraning et al. 2009). MeDIP-seq has been optimized by a combination of photo-cross-linking, exonuclease digestion, and restriction enzyme digestion to achieve near single-nucleotide resolution of 6mA (Chen et al. 2015; Fu et al. 2015). MeDIP-seq, however, is dependent on the antibody specifically recognizing 6mA. Alternative techniques have also been developed to identify where throughout the genome 6mA occurs. One such technique consists of radioactive methylation of DNA followed by restriction digest, electrophoresis, and sequencing (Posfai and Szybalski 1988). Single-molecule real-time sequencing (SMRT-seq) is a next-generation sequencing technique which provides accurate sequence reads and measures the rates of nucleotide incorporation and polymerase pause times during sequencing (Flusberg et al. 2010). Since different DNA modifications result in different kinetic signatures, SMRT-seq can identify every DNA modification at single-base resolution. This technology, however, does have troubles distinguishing several closely related modifications from each other, including 1mA from 6mA. However, when coupled with UHPLC-ms/ms (which can distinguish 1mA from 6mA), this technique can give rather unambiguous confirmation of both the presence and genomic location of 6mA in a specific organism (Greer et al. 2015b). Methylated residues can be confirmed by restriction digest coupled with quantitative PCR to determine the methylation at a specific locus (Fu et al. 2015). Alternatively, sequence-specific probes have been developed that can selectively bind to 6mA or unmodified adenines in specific sequence contexts (Dohno et al. 2010).

To convincingly identify rare modifications, such as 6mA, a combination of multiple complementary techniques is ideal, since each technique has its own set of limitations (Table 1). UHPLC-ms/ms can be complemented by restriction enzyme digestion confirmation (as long as 6mA occurs in the appropriate motif), dot blots and MeDIP with a 6mA-specific antibody, and SMRT-seq.

Table 1 Recent methods for detecting and quantifying 6mA

Detection/Quantitation Method	Description/Limitations	6 mA Specificity	Sensitivity (lower limit of detection)	Genomic sequence information	Reference(s)
6 mA-sensitive restriction enzymes	Restriction endonuclease cleavage of methylated motifs. Cannot detect 6 mA outside of restriction recognition sites.	High (enzyme-dependent)	Can identify single methylated adenine so long as it occurs within specific recognition motifs (e.g. GATC) Preference for hemi-methylated or dually methylated depending on the enzyme	None normally Can be used, in combination with real-time RT PCR, to validate sites identifies by sequencing methods	Bird and Southern (1978), Geier and Modrich (1979)
6mA Dot Blotting	Antibody-dependent semi-quantitative detection of 6 mA levels in genomic DNA samples	High (antibody-dependent) It DNA is not single stranded 6 mA antibodies generally recognize 1 mA as well	Moderate (can not distinguish lowly methylated samples from each other)	None	Achwal et al. (1983)
Immunofluorescence	Antibody-dependent method for detecting 6 mA in whole animals or tissues at cell-level resolution. Very difficult to validate that signal is coming from 6 mA rather than background (ideally need to manipulate the 6 mA regulating enzymes)	High (antibodydependent)	Moderate. Immunofluorescence is not good for assessing relative changes in 6 mA. (antibody-dependent)	None normally Could be used in combination with DNA probes to confirm 6 mA localization in specific genomic regions	Greer et al. (2015 b), Sun et al. (2015)
MeDIP-seq	Antibody-dependent method for identifying genomic regions harboring 6 mA	High (antibody-dependent)	High (antibody- and organism-dependent (C. elegans and D. melanogaster gave very low coverage depth))	Genome-wide 6 mA localization at near base pair resolution	Pomraning et al. (2009), Chen et al. (2015)

SMRT-seq	Provides base modifications at single-nucleotide resolution 6 mA and 1 mA are indistinguishible and quite expensive.	Medium (1 mA and 6 mA are indistinguishible)	Can identify single methylated adenine with sufficient coverage	Single base resolution 6 mA detection genome-wide	Flusberg et al. (2010)
UHPLC-MS/MS	Chemical separation and detection by mass spectrometry.	Highest	High (As low as 0.0001 % 6 mA/A)	None	Yuki et al. (1979), Zhang et al. (2015), Fu et al. (2015), Greer et al. (2015a, b), Huang et al. (2015)
6mA-specific probes	DNA probe containing a formyl group on the O6 position of a G base discriminates between adenine and 6 mA via formation of an interstrand cross-link (ICL). 6 mA can not form ICL. ICLs detected by PAGE or HPLC	Untested for other modifications (such as 1 mA)	Detection by electrophoresis (6 mA has no ICLs and will be single-stranded on the gel)	Can confirm 6 mA in specific genomic locations not a discovery tool	Dohno et al. (2010)
CE-LIF	BODIPY FL EDA binds covalently to the phosphate group of deoxyribonucleotide after activated by carbodiimide reagent. Run by CE-LIF which distinguishes different bases from each other and methylated bases based on migration time.	Can distinguish from 5mC or other bases but untested with 1 mA	Moderate (0.01 % limit)	None	Krais et al. (2010)

Recent methods for detecting and quantifying 6mA are summarized in this table. Relative limitations and sensitivity of each method are discussed and references to the primary papers are cited

Abbreviations: MeDIP-seq methylated DNA immunoprecipitation followed by high-throughput DNA-sequencing, *SMRT-seq* single-molecule real-time sequencing, *UHPLC-ms/ms* ultrahigh-performance liquid chromatography coupled to tandem mass spectrometry. *BODIPY FL EDA* fluorescent dye 4,4-difluoro-5,7-dim ethyl-4-bora-3a,4a-diaza-s-indacene-3-propionyl ethylene diamine hydrochloride, *CE-LIF* capillary electrophoresis with laser-induced fluorescence

6 Enzymes Introducing 6mA

6.1 DNA Methyltransferases

An important step in the confirmation of 6mA as a regulated mark of biological significance has been the identification of enzymes that deposit and remove this mark. It was previously thought that methylated adenines were incorporated premade into genomic DNA. This assumption likely hampered initial efforts to identify 6mA in eukaryotes. A study in the early 1970s concluded that 6mA did not exist in eukaryotes, because radioactively labeled adenines, but not methylated adenines, were incorporated into DNA when added exogenously (Vanyushin et al. 1970). However, several groups demonstrated that DNA could be glycosylated and RNA could be methylated at the N6 position of adenines after incorporation into polynucleotides rather than pre-methylated nucleotides being incorporated during the biosynthesis of polynucleotide (Kornberg et al. 1959, 1961; Fleissner and Borek 1962). These findings led to the hypothesis that methylation occurs after DNA synthesis (Theil and Zamenhof 1963), rather than on unincorporated nucleotides, and spurred attempts to identify the DNA methylating enzymes. The first studies were conducted in *E. coli* by fractionation of total protein lysates followed by methylation assays with each fraction. Early studies identified a single fraction that methylated DNA at the C5 position of cytosines and the N6 position of adenines, but this fraction was only efficient at methylating foreign DNA (Gold et al. 1963; Gold and Hurwitz 1964). Subsequent studies using increasingly subdivided fractions were able to identify multiple adenine and cytosine methyltransferases in *E. coli* (Nikolskaya et al. 1976, 1981).

Additional evidence for the widespread presence and functional importance of 6mA in eukaryotic genomes comes from the observation that members of the MT-A70 family of known or putative N6-adenine methyltransferases exist in most organisms, ranging from bacteria to humans (Luo et al. 2015). Based on structural orthology to other members of the MT-A70 family of methyltransferases, the candidate DNA adenine methyltransferase enzymes in multicellular organisms likely evolved from a bacterial M.MunI-like 6mA methyltransferase, which functions in the host restriction-modification system (Iyer et al. 2011). The MT-A70 family includes both RNA and DNA methyltransferases, including IME4 (also called SPO8) in *S. cerevisiae* (Clancy et al. 2002), DAMT-1 in *C. elegans* (Greer et al. 2015b), and members of the methyltransferase-like (METTL) family in mammals, including METTL3 (an N6-adenosine RNA methyltransferase) (Liu et al. 2014) and METTL4 (a homolog of DAMT-1) (Greer et al. 2015b). Whether the same enzymes catalyze both RNA and DNA adenine methylation in different organisms remains an open question. Notably, biochemical *in vitro* studies have suggested that the mammalian RNA methyltransferase METTL3 also methylates DNA (personal communications C. He), suggesting that the same enzymes are capable of methylating both RNA and DNA in certain contexts, but the substrate specificity (i.e., RNA, DNA, or both) for each member of the different MT-A70 family members remains incompletely characterized. At the structural level, all of these enzymes are

characterized by a 7-ß-strand methyltransferase domain at their C-terminus, fused to a predicted alpha-helical domain at their N-terminus, and require S-adenosyl-L-methionine (AdoMet) as a methyl donor (Iyer et al. 2011). The high degree of amino acid sequence conservation among the predicted N6-adenine methyltransferases motivates further exploration into their potential functional conservation.

How adenine methyltransferases of recently evolved eukaryotes recognize their substrates still remains to be determined. The utilization of adenine methylation by restriction-modification systems suggests that bacterial 6mA methyltransferases evolved to recognize specific sequences for methylation. In bacteria and the unicellular eukaryote *Tetrahymena*, DNA adenine methylation occurs in a palindromic sequence-specific manner in vitro and in vivo (Geier and Modrich 1979; Zelinkova et al. 1990; Bromberg et al. 1982). However, sequence-specific adenine methylation is not observed in all organisms, and some bacterial DNA adenine methyltransferases do not show sequence specificity (Drozdz et al. 2012). Similarly, 6mA sites in *C. elegans* only appeared modestly enriched in specific sequence contexts (Greer et al. 2015b), suggesting that targeted adenines might be selected by more complicated metrics than simple sequence codes. It remains to be seen whether other multicellular eukaryotes, which possess 6mA, show a sequence-specific pattern of adenine methylation (similar to bacteria and unicellular eukaryotes) or whether these organisms show little to no sequence specificity in their adenine methylation pattern, as observed for *C. elegans*. It remains to be seen whether methyltransferases that do not recognize specific DNA sequences are recruited to specific locations of the genome by other DNA-binding proteins or other epigenetic chromatin features.

6.2 Mechanism of 6mA Methyltransferases

Substantial work in prokaryotes has identified the mechanism of action, the preferred methyl donor, and the kinetics of 6mA methyltransferases. Whether these regulatory principles are conserved in eukaryotes remains to be seen. There was an initial debate as to whether N6 was directly methylated or if adenines were first methylated on the N1 position and then; following a Dimroth rearrangement, the methyl group would be transferred to the N6 position. However, the enzyme *EcoRI* had been shown to methylate N6 directly rather than through an initial N1 methylation (Pogolotti et al. 1988). This result, combined with the slow rate of Dimroth reactions at endogenous pH (Macon and Wolfenden 1968), suggests that N6 is the direct target of methyltransferases. This conclusion has been confirmed by the structures of different adenine-N6 methyltransferases in complex with DNA, showing a direct approximation of the N6 atom toward the methyl donor (Goedecke et al. 2001; Horton et al. 2005, 2006).

Early reports identifying that DNA was methylated suggested that AdoMet was the primary methyl donor (Gold et al. 1963), and future work has shown that this cofactor is the predominant methyl donor for not only DNA and RNA methylation but also for methylation of proteins and lipids (Chiang et al. 1996). However,

5,10-methylene tetrahydrofolate has been identified as the methyl donor for tRNAs in *Streptococcus faecalis* and *Bacillus subtilis* (Delk and Rabinowitz 1975; Delk et al. 1976; Urbonavicius et al. 2005). While the enzyme that utilizes 5,10-methylene tetrahydrofolate in *B. subtilis*, *GidA*, is absent in eukaryotes (Urbonavicius et al. 2005), this finding raises the possibility that some DNA methyltransferases might use alternative methyl donors.

DNA methylation rates have been measured for the T4 bacteriophage DNA adenine methyltransferase, T4Dam (Malygin et al. 2000), and the EcoRI adenine methyltransferase (Reich and Mashhoon 1991). For T4Dam the methylation rate constant (k_{meth}) was significantly faster than the overall reaction rate constant (k_{cat}) (0.56 and 0.47 s^{-1} vs. 0.023 s^{-1}) suggesting that product dissociation is the rate-limiting step. Similar but faster results were observed with EcoRI (Reich and Mashhoon 1991). These enzymes function by binding, flipping out the adenine, methylating, and restacking of the modified base (Allan et al. 1998). Whether this holds true for M.MunI-like methyltransferases remains to be determined. Reducing the double-stranded duplex stability did not alter the k_{meth}, suggesting that base flipping is not a rate-limiting step in the methylation reaction (Malygin et al. 2000). Additionally, EcoRI enzyme-DNA complexes were less efficient compared to enzyme-AdoMet complexes, suggesting that the enzyme first binds to AdoMet before methylating its substrates (Reich and Mashhoon 1991). This is opposite to what has been observed with EcoDam and the bacterial 5mC methyltransferase HhaI, where the methyltransferase first binds to DNA, followed by AdoMet (Urig et al. 2002; Wu and Santi 1987), suggesting that the sequence of binding events in the DNA methylation reaction is enzyme dependent.

An important step for the confirmation of the presence and role of 6mA in more recently evolved eukaryotes will be the identification of genuine 6mA methyltransferases. The conservation of MT-A70 domain containing proteins in conjunction with the identification of 6mA in many eukaryotes suggests that this modification is conserved. Whether eukaryotic DNA methyltransferases function in a similar manner to prokaryote methyltransferases remains to be seen. Interestingly, the RNA m6A methyltransferase, METTL3, functions in complex with METTL14 (Liu et al. 2014), raising the possibility that DNA methyltransferase enzymes, like many other chromatin regulating enzymes, function in multi-protein complexes. These multi-protein complexes could help the enzymes achieve their specificity.

7 DNA Adenine Demethylation

The identification of the enzymes that catalyze the removal of 6mA from DNA strongly suggests that 6mA is a regulated and dynamic epigenetic mark. Examination of the enzymes responsible for the removal of DNA base damage fostered the identification and characterization of the DNA demethylation processes. DNA base damage, in the form of 1mA and 3mC, was shown to be removed by the Fe(II)- and α-ketoglutarate-dependent dioxygenase AlkB in *E. coli* (Trewick et al. 2002). The AlkB family of dealkylating enzymes is highly conserved from bacteria to humans

(Fedeles et al. 2015; Wei et al. 1996). AlkB enzymes can demethylate many DNA substrates, including the DNA lesions 1mA, 3mC, and 3mT (Chen et al. 2015; Kamat et al. 2011). Notably, humans have nine AlkB family members (ALKBH1-8 and FTO). Like *E. coli* AlkB enzymes, the mammalian enzymes ALKBH2 and ALKBH3 function in the repair of DNA alkylation damage (Duncan et al. 2002). In addition to their DNA demethylase activity, AlkB members catalyze oxidative demethylation of RNA (Aas et al. 2003). Interestingly, AlkB enzymes in RNA viruses preferentially demethylate RNA substrates, suggesting these AlkBs are necessary for maintaining the integrity of the viral RNA genome (van den Born et al. 2008). More recently, it was found that AlkB family members function in the oxidative demethylation of N6-methyladenosine in RNA, catalyzed by ALKBH5 and FTO in mammals (Jia et al. 2011; Zheng et al. 2013), and that the AlkB family member NMAD-1 in *C. elegans* demethylates 6mA in DNA (Marinus and Lobner-Oleson 2014). FTO was also shown to demethylate 6mA in single-stranded DNA in vitro (Jia et al. 2011), raising the possibility that these enzymes might regulate both DNA and RNA 6mA. Whether NMAD-1 can also demethylate m6A on RNA remains to be tested. Recently ALKBH1 was also shown to demethylate 6mA in single-stranded DNA in vitro (Wu et al. 2016). Additionally ALKBH1 knockout causes an increase in global 6mA levels in mouse embryonic stem cells, and this increase can be rescued by a wild-type but not a catalytic domain mutant of ALKBH1 (Wu et al. 2016), suggesting that ALKBH1 functions as a 6mA demethylase in mammals.

Several studies have begun to dissect the mechanism of action of AlkB demethylases. In the presence of their essential cofactors α-ketoglutarate and Fe(II), AlkB demethylases use molecular oxygen to oxidize the methyl group of 6mA, forming the unstable intermediate 6-hydroxymethyladenine (6hmA), which spontaneously releases formaldehyde, regenerating the unmodified adenine base (Fig. 2) (Fedeles et al. 2015). Whether the same mechanism holds true for the demethylation of 6mA in eukaryotes and, if so, whether 6hmA has any additional function remains to be seen. In mammals, FTO was recently shown to oxidize m6A on RNA to N6-hydroxymethyladenosine (hm6A) and N6-formyladenosine (f6A) (Fu et al. 2013). These mRNA derivatives have half-lives of ~3 h (Fu et al. 2013), suggesting that if 6hmA does have additional functions, they would require a 6hmA specific binding protein which could stabilize the intermediate. The same oxidation reaction mechanism is used by AlkB enzymes to demethylate 1mA and 3mC during the cellular response to DNA alkylation damage (Falnes et al. 2002; Trewick et al. 2002).

In addition to demethylation of 6mA by the AlkB demethylase family, 6mA can also be converted to hypoxanthine by a 6mA deaminase (Kamat et al. 2011). This modified base can then undergo base excision repair by hypoxanthine DNA glycosylases of the AlkA family (Saparbaev and Laval 1994) (Fig. 2). If hypoxanthine is not removed, it can cause a transition mutation (an AT pair would be converted to a GC pair), since hypoxanthine pairs with cytosine instead of thymine. Recently, 6mA was found to be correlated with increased point mutations in *Neisseria meningitidis* (Sater et al. 2015), suggesting that this modified base might be mutagenic, potentially as a consequence of unrepaired 6mA

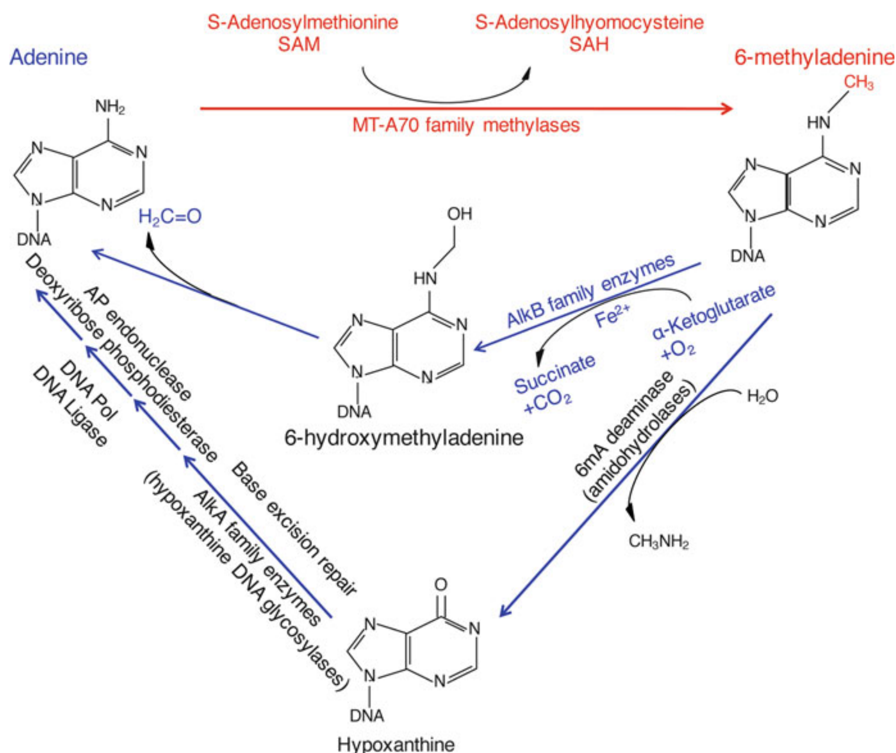


Fig. 2 Mechanisms of N6-adenine methylation and demethylation. MT-A70 family methylases catalyze the methylation of adenine at the sixth position of the purine ring. MT-A70 methylases use S-adenosylmethionine (SAM) as their methyl donor to generate 6-methyladenine and S-adenosylhomocysteine (SAH). Adenine could be regenerated from 6mA by several different enzymatic mechanisms: AlkB family enzymes catalyze the oxidative demethylation of 6mA. AlkB enzymes require α-ketoglutarate and Fe^{2+} and use oxygen to oxidize the methyl group. This oxidative demethylation reaction first generates 6-hydroxymethyladenine, which releases its formaldehyde group to generate adenine. Alternatively, 6mA can be deaminated and subsequently removed via the base excision repair pathway. First, 6mA deaminase hydrolyzes the methylamine to generate hypoxanthine. Hypoxanthine is recognized as a damaged base by AlkA family enzymes, which cleave the glycosyl bond to remove the base. Apurinic (AP) endonuclease cleaves the phosphodiester backbone at the abasic site, exposing the residual 5′ deoxyribose phosphate group, which is then removed by deoxyribose phosphodiesterase. Finally, DNA polymerase I incorporates the unmodified adenine, and DNA ligase catalyzes the formation of the phosphodiester bond

deamination. However, 6mA deaminases in *Neisseria meningitidis* have not yet been identified. Similarly to 6mA deamination, 5mC is converted to thymine when deaminated, which leads to a transition mutation (Lindahl and Nyberg 1974; Heindell et al. 1978). Deamination of adenine, 6mA, or cytosine all leads to non-natural bases, which can easily be identified by specific glycosylases. Deamination of 5mC, in contrast, leads to thymine, which requires a more complicated repair process. This more direct mutational path might explain why 5mC is more prone to mutation than 6mA. This divergence begs the question as to why evolution has

selected for a higher prevalence of the more mutagenic DNA modification in more recently evolved species.

In *E. coli*, AlkB expression is induced by DNA damage and the enzyme functions in DNA repair via direct removal of base alkylation damage (Trewick et al. 2002). *AlkB* mutant *E. coli* are sensitized to cell death induced by the alkylating agent methyl methanesulfonate (MMS), and the predicted human ortholog of AlkB is sufficient to partially rescue the MMS-induced cytotoxicity seen in AlkB mutants (Wei et al. 1996). Interestingly, MMS treatment of human skin fibroblasts did not result in the same induction of AlkB seen in *E. coli*, suggesting that the regulation of AlkB expression may have diverged during the evolution of more recent eukaryotes (Wei et al. 1996) or that the induction by different alkylating agents is cell-type specific and may only occur in certain cell types. In *C. elegans*, deletion of NMAD-1, a member of the AlkB family, causes a global increase in 6mA, and purified NMAD-1 is capable of demethylating N6-adenine methylated oligonucleotides *in vitro* (Greer et al. 2015b). Importantly, mutation of the NMAD-1 catalytic domain abolished the *in vitro* demethylase activity of NMAD-1, identifying NMAD-1 as a 6mA demethylase in *C. elegans* and highlighting the mechanistic conservation of AlkB enzymes from bacteria to metazoa (Greer et al. 2015b).

Interestingly, a different family of enzymes, the so-called ten-eleven translocation (Tet) proteins, has been shown to demethylate 5mC in many organisms (Tahiliani et al. 2009; Ito et al. 2010, 2011). Unlike AlkB proteins, whose crystal structures have revealed that the enzymes flip out the base to facilitate demethylation (Yang et al. 2008; Sundheim et al. 2008), crystal structures of the TET family catalytic domains are not suitable for accommodating flipped out purines (Aravind et al. 2015) suggesting that they cannot act on mA. Moreover, the TET family has a good phyletic correlation with DNA cytosine methylases, but not with DAMT-1 or other Dam family methylases (Aravind et al. 2015). Additionally in bacteria, there is little evidence that TET is capable of demethylating purines (Aravind et al. 2015). Given these findings, it is surprising that the *D. melanogaster* ortholog of Tet (named DMAD) was reported to function as a 6mA demethylase on DNA (Zhang et al. 2015). Nuclear extracts from DMAD mutant flies showed reduced *in vitro* demethylation activity compared to nuclear extracts from wild-type flies, while addition of purified DMAD was sufficient to increase adenine demethylation in these assays (Zhang et al. 2015). It remains to be seen whether this 6mA demethylase activity can be biochemically confirmed using purified DMAD and whether Tet proteins play a role as 6mA demethylases as well.

8 Proteins Binding 6mA

Beyond the machinery that catalyzes addition and removal of 6mA, cells have evolved mechanisms to recognize 6mA as a regulatory signal that can be translated into different biological consequences (see below, Sect. 9). We will discuss later in this chapter the direct chemical consequences of adenine methylation, but 6mA can be recognized by specific effector molecules that alter chromatin architecture and/or transcriptional

states directly, or indirectly, via recruitment of other DNA-binding proteins. Alternatively, methylation could function by preventing binding of chromatin factors. Methyladenine-binding proteins have evolved to recognize and transduce 6mA signals into specific biological outcomes. For example, in *E. coli* the MutS enzyme binds to mismatch base pairs as a homodimer, facilitating recruitment of the MutL protein, which binds MutS. The MutS-MutL-DNA complex then loops DNA until it finds the nearest hemimethylated GATC site, which is bound by the endonuclease MutH. Upon binding of MutL-MutS to the MutH-DNA complex, MutH is activated and nicks the unmethylated daughter strand, allowing helicase and exonucleases to excise the single-stranded mismatch region (Su and Modrich 1986). Thus, hemimethylated of GATC sites are used to specifically direct mismatch repair of the daughter strand (Lahue et al. 1987) Similarly, the *oriC* region of *E. coli* is hemimethylated to prevent premature replication before the cell has divided. These hemimethylated adenine sites are recognized and bound by the SeqA protein (Brendler et al. 1995; Slater et al. 1995), which prevents assembly of the DNA replication machinery at this region (von Freiesleben et al. 1994; Wold et al. 1998). The crystal structure for SeqA has revealed why SeqA binds preferentially to hemimethylated over fully methylated DNA (Guarne et al. 2002; Fujikawa et al. 2004), highlighting the importance of determining the crystal structure of 6mA-binding proteins for deciphering the chemical and biological consequences of their binding. Thus far, these binding proteins have only been identified in prokaryotes, but an important next step to fully understand the possible biological roles of 6mA will be to identify eukaryotic 6mA-binding proteins.

9 Biological Functions of 6mA

The direct effects of adenine methylation on the structure of DNA and its roles in prokaryote biology have been well characterized. Whether 6mA plays a conserved functional role in eukaryotes remains to be seen, but discussing its functional effects in prokaryotes raises several interesting potential functions which will need to be further explored in eukaryotes.

9.1 Effects of Adenine Methylation on DNA Structure

One possible role for adenine methylation, beyond providing a binding site for effector proteins, is to directly alter the overall structure of DNA. An early crystal structure suggested that 6mA might alter the secondary structure of DNA (Sternglanz and Bugg 1973). Adenine methylation is thought to affect DNA double helix formation through altering both base pair stability and base stacking. Ultraviolet photoelectron studies suggested that adenine methylation would lower the ionization potentials and cause the destabilization of valence electrons to increase base stacking in methylated adenines (Peng et al. 1976). This increased base stacking would be offset by a slight destabilization of base pairing ranging from ~0.35 to 0.95 kcal/mol (Engel and von Hippel 1978b). Interestingly, 5mC behaves oppositely to 6mA in these regards. 5mC causes an increase in helix stability, while adenine methylation destabilizes the DNA,

as measured by denaturing gradient gel electrophoresis (Collins and Myers 1987). 6mA within GATC sequences causes slight DNA unwinding of 0.5°/methyl group (Cheng et al. 1985). Consistent with these observations, two-dimensional NMR studies suggest that, in almost all cases, 6mA has only minor effects on helix conformation, as it retains the canonical B-form (Fazakerley et al. 1985, 1987; Quignard et al. 1985). 6mA occurring directly after thymines, on the other hand, causes severe unwinding and bending of the DNA helix relative to the canonical B conformation (Fazakerley et al. 1989). However, 6mA lowers melting temperatures and slows the rate of helix formation, as demonstrated by enthalpy of dissociation studies (Quignard et al. 1985; Fazakerley et al. 1985). These studies suggest that methylated adenines are associated with DNA regions that spend prolonged periods in the open state. These effects were confirmed by cruciform extrusion assays where 5mC inhibits extrusion and 6mA facilitates initial opening of DNA (Murchie and Lilley 1989). These consequences seem to be in line with the reported effects of 5mC and 6mA on gene transcription; 5mC is generally believed to be a repressor of gene transcription when it occurs at promoters, while 6mA is hypothesized to be an activator. However, the correlation between 5mC and gene transcription is dependent on the genomic context in which it occurs. When 5mC occurs within gene bodies, rather than promoters, it is correlated with gene transcription (reviewed in (Jones 2012)). Thus, the effects of 6mA on gene transcription may depend on its location in the genome as well.

The effects of 6mA on the thermodynamic stability and folding of DNA appear to be sequence specific (Fazakerley et al. 1987). Indeed, when 6mA occurs directly after a T, this can cause a highly altered structure that is overwound and bent (Fazakerley et al. 1989). While 6mA does not dramatically alter helix rigidity (Hagerman and Hagerman 1996; Mills and Hagerman 2004), it can increase DNA curvature to variable degrees, depending on sequence context (Diekmann 1987).

9.2 Restriction-Modification Systems

In prokaryotes, DNA N6-adenine methylation is used to discriminate self from foreign DNA, as part of restriction-modification systems, a bacterial immune system by which pathogenic DNA from bacteriophages is recognized by endonucleases that selectively cleave unmethylated DNA at specific restriction sites that are methylated in the host's genome and thus protected from endonuclease digestion (Low et al. 2001; Iyer et al. 2011). Interestingly, enterobacteriophages appear to have evolved to contain fewer GATCs to avoid the GATC R-M system of their hosts (McClelland 1984). This system does not appear to be conserved in eukaryotes that have evolved more complex immune systems.

9.3 DNA Damage Control

Early reports suggested that *dam* mutant *E. coli* had higher mutation rates and were more sensitive to UV and mitomycin C, suggesting that 6mA could protect against DNA damage (Marinus and Morris 1974). It was subsequently suggested that 6mA

could help to distinguish the parental DNA strand from the mutated daughter strand (Glickman et al. 1978; Glickman 1979). Similarly, *Penicillium chrysogenum* mutants deficient in 6mA had higher sensitivity to mutagenic agents without changes in the number of mutations (Rogers et al. 1986).

In *E. coli* and other gram-negative bacteria, DNA adenine methylation plays an important role in the DNA mismatch repair pathway, a strand-specific repair pathway that relies on the transient post-replicative hemimethylation of DNA. The DNA adenine methylase, Dam, binds to hemimethylated DNA substrates and methylates GATC sites after DNA replication. The delay between DNA synthesis and methylation of the newly synthesized daughter strand is crucial for the fidelity of DNA mismatch repair (Pukkila et al. 1983). When DNA replication errors lead to base pair mismatches, the DNA repair machinery uses adenine methylation to distinguish the already methylated template strand from the newly synthesized unmethylated daughter strand. As described above (6mA-binding proteins) hemimethylated DNA allows MutL, MutS, and MutH to identify and specifically cleave the daughter strand, allowing helicase and exonucleases to excise the single-stranded mismatch region. Subsequently, DNA polymerase III resynthesizes the mismatch region of single-stranded DNA using the methylated parental strand as a template (Pukkila et al. 1983).

9.4 Effect on Transcription

Several studies have suggested that N6-adenine methylation correlates with increased gene expression. Whether this is due to the direct effect on relaxing DNA structure (as discussed above), recruitment of 6mA-specific binding proteins, or both remains unknown. It is still also unclear as to whether this phenomenon is conserved across all organisms that contain 6mA. While 5mC CpG methylation had little effect on transcription in barley, 6mA methylation increased transcription two- to fivefold (Rogers and Rogers 1995). Similarly, 6mA but not 5mC methylation increased gene expression by 3–50-fold of reporter constructs in tobacco or wheat protoplast or intact wheat tissues (Graham and Larkin 1995). Luciferase reporter constructs purified from dam + dcm + bacteria (with 5mC and 6mA methylation) had two- to sixfold increased luciferase production compared to constructs purified from dam-dcm-bacteria in rat or mouse cell lines or when electroporated into mice (Allamane et al. 2000). Together, these results suggest that 6mA promotes gene expression.

6mA can also directly affect binding of transcription factors. Methylation of a HNF1 binding site reduces HNF1 binding affinity, but this only causes a minor reduction in gene transcription (Tronche et al. 1989; Lichtsteiner and Schibler 1989). Conversely, 6mA increases binding affinity for the transcription factor AGP1 in tobacco (Sugimoto et al. 2003). These results suggest that the effects of adenine methylation on transcription will be sequence and transcription factor specific.

Similar to DNA cytosine methylation in metazoa, bacterial DNA adenine methylation regulates gene expression programs, including those related to virulence and phase variation (Low et al. 2001; Wallecha et al. 2002; Zaleski et al. 2005;

Sarnacki et al. 2013), suggesting that 6mA levels might be sensitive to changes in environmental conditions. Similarly, recent data suggest that 6mA may play a role in transcriptional regulation in the single-celled eukaryote *Chlamydomonas reinhardtii*, where 6mA occurs preferentially near actively transcribed genes (Fu et al. 2015). As preliminary evidence that 6mA levels might be relevant to human physiology and disease, it was reported that human patients with type 2 diabetes have reduced levels of m6A on RNA and 6mA on DNA, as measured by HPLC-ms/ms, and it was proposed that these differences might be regulated by the cellular fat mass and obesity-associated protein (FTO) (Huang et al. 2015), which was shown to function as an RNA m6A and single-stranded DNA 6mA demethylase (Jia et al. 2011) and DNA 3mT demethylase (Gerken et al. 2007). Future studies will be required to confirm the existence of 6mA in human DNA using independent detection methods. Despite recent progress in defining the potential functions of 6mA in different organisms, the roles of 6mA in more recently evolved eukaryotes, including its possible roles in human health and disease, remain unknown. Given the high degree of evolutionary conservation of MT-A70 family methyltransferases and Alkb family demethylases, along with the recent discovery of 6mA in eukaryotes, we propose that 6mA is likely to play an important role in the regulation of diverse biological processes in metazoa.

9.5 Nucleosome Positioning

In the protists *Tetrahymena* and *Chlamydomonas*, 6mA is preferentially located in the linker regions between nucleosomes (Karrer and VanNuland 2002; Fu et al. 2015; Pratt and Hattman 1983), raising the possibility that 6mA could help to direct nucleosome positioning. Alternatively, enrichment of 6mA in linker regions may reflect increased accessibility or recruitment of the methyltransferase at regions of open chromatin. Enrichment for 6mA in specific genomic regions was not observed in *C. elegans* (Greer et al. 2015b), but the analysis was performed on mixed tissue samples, which could have obscured any positional bias that may exist in specific cell types. In future studies, it will be interesting to examine whether 6mA directs nucleosome positioning and whether it does so in a conserved manner or whether other open chromatin modifications can direct N6-adenine methylation at those sites.

9.6 Cell Cycle Regulation

N6-adenine methylation marks regions for DNA replication initiation in prokaryotes and has been shown to alter the rate of cell cycle progression. In *E. coli*, the Dam methyltransferase is necessary for precise timing between DNA replication events (Bakker and Smith 1989; Boye and Lobner-Olesen 1990). The hemimethylation of DNA plays an important role in modulating the initiation of DNA replication; the SeqA protein binds to hemimethylated DNA adjacent to the origin of replication *OriC*, preventing its methylation by Dam and leading to a delay in DNA

replication before the cell has divided, which is only initiated from a fully methylated promoter (Low et al. 2001; Lu et al. 1994). When DNA replication is desired, adenine methylation at the oriC region lowers the thermal melting temperature which could facilitate the unwinding at the origin of replication (Yamaki et al. 1988). Interestingly, 6mA also slows the rate of DNA polymerase I catalysis, presumably due to the effects of 6mA on base pairing (discussed above) (Engel and von Hippel 1978a).

In *Caulobacter crescentus*, the cell cycle-regulated DNA adenine methylase (CcrM) controls the timing of DNA replication and progression through the cell cycle (Collier et al. 2007). In contrast to the *E. coli* Dam methylase, which does not have a preference for hemimethylated sites, *C. crescentus* CcrM preferentially methylates hemimethylated DNA after replication (Berdis et al. 1998) and is essential for cell viability (Stephens et al. 1996). In *C. crescentus*, 6mA levels change throughout the cell cycle from fully to hemimethylated as the replication forks progress (Kozdon et al. 2013). The promoter of the replication initiation factor DnaA is preferentially activated when its promoter is fully methylated, leading to DnaA accumulation and progression through the cell cycle (Collier et al. 2007). Whether 6mA plays a similar role in controlling the cell cycle in eukaryotes remains to be seen.

9.7 Transgenerational Inheritance

DNA methylation at palindromic sites provides the most parsimonious method by which epigenetic information could be transmitted across generations. Because of the semiconservative nature of DNA replication, methylation events on the parental strand can be replicated on the newly synthesized daughter strand. In mammals, 5mC methylation patterns are established by the de novo methyltransferases Dnmt3a and Dnmt3b during early embryonic development (Okano et al. 1999). Inheritance of cytosine methylation patterns through cell division is mediated by the maintenance methyltransferase Dnmt1 (Bestor et al. 1988). Dnmt1 preferentially binds hemimethylated DNA at the replication fork and copies parental-strand methylation patterns onto the unmethylated daughter strand (Stein et al. 1982; Yoder et al. 1997; Bestor 2000; Martin and Zhang 2007). Whether adenine methylation propagates nongenetic information through cell divisions or from parents to their offspring remains to be seen. However, there are some hints that 6mA could transmit nongenetic information. Labeling experiments showed that newly synthesized *E. coli* DNA in Okazaki fragments were quickly N6-adenine methylated (Marinus 1976), consistent with the idea that parental methylation patterns might be passed on to their descendants during DNA replication. In some bacteria, DNA adenine methylation is tightly coordinated with cell division Casadesus and Low (2006)) (see Sect. 9.6 above), enabling inheritance of parental methylation patterns. Thus, a key unanswered question is whether there exists a mode of inheritance of adenine methylation in eukaryotes or whether different organisms have evolved different mechanisms for the inheritance of parental DNA methylation through somatic nuclear

divisions and across generations. In the ciliate *Tetrahymena thermophila* macronucleus, analysis of methylation patterns using methylation-sensitive restriction enzymes showed that both actively replicating and nonreplicating DNA contained hemimethylated sites and that the vegetatively growing macronucleus contained a combination of partially methylated sites and fully methylated sites (Capowski et al. 1989). These findings are inconsistent with a simple semiconservative 6mA inheritance mechanism and suggest that inheritance of 6mA in some organisms may rely on hemimethylation-independent mechanisms of 6mA maintenance through cell division (Capowski et al. 1989).

In *C. elegans*, loss of the histone H3 lysine 4 dimethyl (H3K4me2) demethylase *spr-5* causes a progressive transgenerational loss of fertility (Katz et al. 2009) and a transgenerational extension in life span (Greer et al. 2016). This is accompanied by a progressive decline in H3K9me3 and accumulation of H3K4me2 and 6mA (Greer et al. 2014, 2015b). Deletion of the 6mA demethylase, *nmad-1*, accelerates the progressive fertility decline, while deletion of the putative 6mA methyltransferase, *damt-1*, suppresses the transgenerational H3K4me2 accumulation, fertility, and longevity phenotypes (Greer et al. 2015b; Greer et al. 2016), raising the possibility that 6mA might transmit epigenetic information across generations. It remains to be seen whether methylated adenines themselves are transmitted across generations as they are transmitted across cell divisions, or whether 6mA is erased in the germ line and established de novo during somatic development (see Sect. 9.6), or if 6mA is more indirectly involved in these processes. Future studies will also reveal whether 6mA can regulate transgenerational inheritance in other species.

Many years of research have shown that chromatin modifications do not occur in isolation but rather actively communicate with each other. For example, 5mC and H3K9me3 are coordinately regulated in plants. The H3K9 methyltransferase binds to 5mC methylated DNA (Jackson et al. 2002; Johnson et al. 2007; Malagnac et al. 2002), and the DNA methyltransferase binds to H3K9me-containing nucleosomes (Du et al. 2012). It is possible that a similar reciprocal cross talk occurs between 6mA and H3K4 methylation in *C. elegans*, as described in the previous paragraph (Greer et al. 2015b). It remains to be seen whether this reciprocal cross talk is real and whether other species show a similar co-association between 6mA levels and H3Kme2 levels. Future work should reveal whether 6mA methyltransferases can bind to specific methylated histones to direct DNA methylation to particular loci.

10 Conclusions and Future Directions

As detection techniques are becoming increasingly sensitive, 6mA has begun to be convincingly observed in several metazoa. The conservation of 6mA methyltransferases and demethylases along with the initial detection of 6mA in several metazoa suggests that N6-adenine methylation might be a conserved signaling modification. However, it will be important to rigorously examine whether 6mA is present across the tree of life using a combination of rapidly evolving detection techniques (discussed in this review). For metazoa that are confirmed to have 6mA in their DNA, it

will be important to define the biological functions of 6mA and its genomic localization patterns in different cell types. A fundamental question is whether the biological functions of 6mA in bacteria are conserved in higher eukaryotes or whether 6mA has evolved new biological functions in these organisms. As 6mA occurs less frequently in more recently evolved organisms, this might reflect a more specialized functional role.

A growing body of work has revealed an important role for m6A on mRNAs in the regulation of gene expression and cellular differentiation in eukaryotes (Niu et al. 2013; Meyer et al. 2012; Dominissini et al. 2012; Deng et al. 2015; Wang et al. 2014, 2015; Zhou et al. 2015; Yue et al. 2015; Batista et al. 2014). Therefore, another open question is whether N6-adenine methylation of DNA is coordinately regulated with N6-adenine methylation on RNA. Given that substrates of the AlkB family of demethylases and MT-A70 family of methyltransferases can include both RNA and DNA, it will be of interest to better characterize the substrate specificity of these enzymes in different organisms and to examine whether the same enzymes regulate both RNA and DNA N6-adenine methylation in different organisms. Moreover, it will be relevant to find out if in cases of overlapping substrate specificities, whether methylation of DNA or RNA (or both) is the biologically relevant signal under different physiological conditions.

The inheritance of 6mA methylation during bacterial cell division (Wion and Casadesus 2006) raises the question of whether 6mA can be inherited in eukaryotes. Is 6mA passed on through successive generations or erased in the germ line? Recently described paradigms of transgenerational inheritance in *C. elegans* have raised the possibility that 6mA itself might carry epigenetic information across generations (Greer et al. 2015b). Alternatively, 6mA might communicate with other heritable epigenetic marks that reciprocally regulate the levels of 6mA. Future studies should reveal whether 6mA is incompletely erased in the germ line and inherited in subsequent generations. In mice, 5mC is mostly erased by passive demethylation during the expansion of primordial germ cells preceding the formation of gametes (Seisenberger et al. 2012); methylation patterns are then reestablished during early embryonic development by the de novo methyltransferases Dnmt3a and Dnmt3b (Okano et al. 1999). However, some regions of 5mC, such as those near imprinted genes, escape the typical erasure and can therefore carry nongenetic information across generations (Breiling and Lyko 2015). Whether a similar situation exists for 6mA remains to be seen.

Given the dynamic nature of 5mC in mammalian development and cell differentiation (Okano et al. 1999), it will be of interest to define the dynamics and potential functions of 6mA during mammalian development, if its presence in mammals can be rigorously confirmed. Future studies should also reveal the environmental factors that regulate the levels of 6mA and its modifying enzymes in metazoa, which should provide clues to its evolutionary conservation and biological relevance. The diversity of methods for detection of 6mA in DNA will allow for comprehensive and detailed examination of 6mA's presence, localization patterns, and potential functions in the genomes of diverse organisms. All in all, the newly developed and more sensitive tools for detection, along with the recent discovery of 6mA in metazoa, open an exciting new chapter of discovery in the field of adenine methylation.

Acknowledgements We thank S. Burger, N. O'Brown, and E. Pollina for critical reading of the manuscript. We thank C. He for helpful discussions. The work from the Greer laboratory is supported by a grant from the NIH (AG043550). Z.K.O. is supported by 5T32HD7466-19.

References

Aas PA, Otterlei M, Falnes PO, Vagbo CB, Skorpen F, Akbari M, et al. Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. Nature. 2003;421(6925):859–63. doi:10.1038/nature01363.

Achwal CW, Iyer CA, Chandra HS. Immunochemical evidence for the presence of 5mC, 6mA and 7mG in human, Drosophila and mealybug DNA. FEBS Lett. 1983;158(2):353–8.

Adams RL, McKay EL, Craig LM, Burdon RH. Methylation of mosquito DNA. Biochim Biophys Acta. 1979;563(1):72–81.

Allamane S, Jourdes P, Ratel D, Vicat JM, Dupre I, Laine M, et al. Bacterial DNA methylation and gene transfer efficiency. Biochem Biophys Res Commun. 2000;276(3):1261–4. doi:10.1006/bbrc.2000.3603.

Allan BW, Beechem JM, Lindstrom WM, Reich NO. Direct real time observation of base flipping by the EcoRI DNA methyltransferase. J Biol Chem. 1998;273(4):2368–73.

Ammermann D, Steinbruck G, Baur R, Wohlert H. Methylated bases in the DNA of the ciliate Stylonychia mytilus. Eur J Cell Biol. 1981;24(1):154–6.

Aravind L, Zhang D, Iyer LM. The TET/JBP family of nucleic acid base-modifying 2-oxoglutarate and iron-dependent dioxygenases. In: Hausinger R, Schofield C, editors. 2-oxoglutarate-dependent oxygenases. Royal Society of Chemistry; 2015;3(12);289.

Babinger P, Kobl I, Mages W, Schmitt R. A link between DNA methylation and epigenetic silencing in transgenic Volvox carteri. Nucleic Acids Res. 2001;29(6):1261–71.

Bakker A, Smith DW. Methylation of GATC sites is required for precise timing between rounds of DNA replication in Escherichia coli. J Bacteriol. 1989;171(10):5738–42.

Batista PJ, Molinie B, Wang J, Qu K, Zhang J, Li L, et al. m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. Cell Stem Cell. 2014;15(6):707–19. doi:10.1016/j.stem.2014.09.019.

Berdis AJ, Lee I, Coward JK, Stephens C, Wright R, Shapiro L, et al. A cell cycle-regulated adenine DNA methyltransferase from Caulobacter crescentus processively methylates GANTC sites on hemimethylated DNA. Proc Natl Acad Sci U S A. 1998;95(6):2874–9.

Bestor TH. The DNA, methyltransferases of mammals. Hum Mol Genet. 2000;9(16):2395–402.

Bestor T, Laudano A, Mattaliano R, Ingram V. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. J Mol Biol. 1988;203(4):971–83.

Bird A. DNA methylation patterns and epigenetic memory. Genes Dev. 2002;16(1):6–21. doi:10.1101/gad.947102.

Bird AP, Southern EM. Use of restriction enzymes to study eukaryotic DNA methylation: I. The methylation pattern in ribosomal DNA from Xenopus laevis. J Mol Biol. 1978;118(1):27–47.

Bodi Z, Zhong S, Mehra S, Song J, Graham N, Li H, et al. Adenosine methylation in Arabidopsis mRNA is associated with the 3′ end and reduced levels cause developmental defects. Front Plant Sci. 2012;3:48. doi:10.3389/fpls.2012.00048.

Boye E, Lobner-Olesen A. The role of dam methyltransferase in the control of DNA replication in E. coli. Cell. 1990;62(5):981–9.

Breiling A, Lyko F. Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. Epigenetics Chromatin. 2015;8:24. doi:10.1186/s13072-015-0016-6.

Brendler T, Abeles A, Austin S. A protein that binds to the P1 origin core and the oriC 13mer region in a methylation-specific fashion is the product of the host seqA gene. EMBO J. 1995; 14(16):4083–9.

Bromberg S, Pratt K, Hattman S. Sequence specificity of DNA adenine methylase in the protozoan Tetrahymena thermophila. J Bacteriol. 1982;150(2):993–6.

Campbell JL, Kleckner N. E. coli oriC and the dnaA gene promoter are sequestered from dam methyltransferase following the passage of the chromosomal replication fork. Cell. 1990; 62(5):967–79.

Capowski EE, Wells JM, Harrison GS, Karrer KM. Molecular analysis of N6-methyladenine patterns in Tetrahymena thermophila nuclear DNA. Mol Cell Biol. 1989;9(6):2598–605.

Capuano F, Mulleder M, Kok R, Blom HJ, Ralser M. Cytosine DNA methylation is found in Drosophila melanogaster but absent in Saccharomyces cerevisiae, Schizosaccharomyces pombe, and other yeast species. Anal Chem. 2014;86(8):3697–702. doi:10.1021/ac500447w.

Casadesus J, Low D. Epigenetic gene regulation in the bacterial world. Microbiol Mol Biol Rev. 2006;70(3):830–56. doi:10.1128/MMBR.00016-06.

Chen K, Luo GZ, He C. High-resolution mapping of N(6)-methyladenosine in transcriptome and genome using a photo-crosslinking-assisted strategy. Methods Enzymol. 2015;560:161–85. doi:10.1016/bs.mie.2015.03.012.

Cheng SC, Herman G, Modrich P. Extent of equilibrium perturbation of the DNA helix upon enzymatic methylation of adenine residues. J Biol Chem. 1985;260(1):191–4.

Chiang PK, Gordon RK, Tal J, Zeng GC, Doctor BP, Pardhasaradhi K, et al. S-Adenosylmethionine and methylation. FASEB J. 1996;10(4):471–80.

Clancy MJ, Shambaugh ME, Timpte CS, Bokar JA. Induction of sporulation in Saccharomyces cerevisiae leads to the formation of N6-methyladenosine in mRNA: a potential mechanism for the activity of the IME4 gene. Nucleic Acids Res. 2002;30(20):4509–18.

Collier J, McAdams HH, Shapiro L. A DNA methylation ratchet governs progression through a bacterial cell cycle. Proc Natl Acad Sci U S A. 2007;104(43):17111–6. doi:10.1073/pnas.0708112104.

Collins M, Myers RM. Alterations in DNA helix stability due to base modifications can be evaluated using denaturing gradient gel electrophoresis. J Mol Biol. 1987;198(4):737–44.

Cummings DJ, Tait A, Goddard JM. Methylated bases in DNA from Paramecium aurelia. Biochim Biophys Acta. 1974;374(1):1–11.

Degnen ST, Morris NR. Deoxyribonucleic acid methylation and development in Caulobacter bacteroides. J Bacteriol. 1973;116(1):48–53.

Delk AS, Rabinowitz JC. Biosynthesis of ribosylthymine in the transfer RNA of Streptococcus faecalis: a folate-dependent methylation not involving S-adenosylmethionine. Proc Natl Acad Sci U S A. 1975;72(2):528–30.

Delk AS, Romeo JM, Nagle Jr DP, Rabinowitz JC. Biosynthesis of ribothymidine in the transfer RNA of Streptococcus faecalis and Bacillus subtilis. A methylation of RNA involving 5,10-methylenetetrahydrofolate. J Biol Chem. 1976;251(23):7649–56.

Deng X, Chen K, Luo GZ, Weng X, Ji Q, Zhou T, et al. Widespread occurrence of N6-methyladenosine in bacterial mRNA. Nucleic Acids Res. 2015;43(13):6557–67. doi:10.1093/nar/gkv596.

Diekmann S. DNA methylation can enhance or induce DNA curvature. EMBO J. 1987;6(13):4213–7.

Dohno C, Shibata T, Nakatani K. Discrimination of N6-methyl adenine in a specific DNA sequence. Chem Commun (Camb). 2010;46(30):5530–2. doi:10.1039/c0cc00172d.

Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature. 2012;485(7397):201–6. doi:10.1038/nature11112.

Drozdz M, Piekarowicz A, Bujnicki JM, Radlinska M. Novel non-specific DNA adenine methyltransferases. Nucleic Acids Res. 2012;40(5):2119–30. doi:10.1093/nar/gkr1039.

Du J, Zhong X, Bernatavichute YV, Stroud H, Feng S, Caro E, et al. Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell. 2012;151(1):167–80. doi:10.1016/j.cell.2012.07.034.

Duncan T, Trewick SC, Koivisto P, Bates PA, Lindahl T, Sedgwick B. Reversal of DNA alkylation damage by two human dioxygenases. Proc Natl Acad Sci U S A. 2002;99(26):16660–5. doi:10.1073/pnas.262589799.

Dunn DB, Smith JD. Occurrence of a new base in the deoxyribonucleic acid of a strain of Bacterium coli. Nature. 1955;175(4451):336–7.

Dunn DB, Smith JD. The occurrence of 6-methylaminopurine in deoxyribonucleic acids. Biochem J. 1958;68(4):627–36.

Ehrlich M, Gama-Sosa MA, Huang LH, Midgett RM, Kuo KC, McCune RA, et al. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. Nucleic Acids Res. 1982;10(8):2709–21.

Ehrlich M, Gama-Sosa MA, Carreira LH, Ljungdahl LG, Kuo KC, Gehrke CW. DNA methylation in thermophilic bacteria: N4-methylcytosine, 5-methylcytosine, and N6-methyladenine. Nucleic Acids Res. 1985;13(4):1399–412.

Ehrlich M, Wilson GG, Kuo KC, Gehrke CW. N4-methylcytosine as a minor base in bacterial DNA. J Bacteriol. 1987;169(3):939–43.

Engel JD, von Hippel PH. D(M6ATP) as a probe of the fidelity of base incorporation into polynucleotides by Escherichia coli DNA polymerase I. J Biol Chem. 1978a;253(3):935–9.

Engel JD, von Hippel PH. Effects of methylation on the stability of nucleic acid conformations. Studies at the polymer level. J Biol Chem. 1978b;253(3):927–34.

Falnes PO, Johansen RF, Seeberg E. AlkB-mediated oxidative demethylation reverses DNA damage in Escherichia coli. Nature. 2002;419(6903):178–82. doi:10.1038/nature01048.

Fazakerley GV, Guy A, Teoule R, Quignard E, Guschlbauer W. A proton 2D-NMR study of an oligodeoxyribonucleotide containing N6-methyladenine:d(GGm6ATATCC). Biochimie. 1985; 67(7–8):819–22.

Fazakerley GV, Quignard E, Teoule R, Guy A, Guschlbauer W. A two-dimensional 1H-NMR study of the dam methylase site: comparison between the hemimethylated GATC sequence, its unmethylated analogue and a hemimethylated CATG sequence. The sequence dependence of methylation upon base-pair lifetimes. Eur J Biochem. 1987;167(3):397–404.

Fazakerley GV, Gabarro-Arpa J, Lebret M, Guy A, Guschlbauer W. The GTm6AC sequence is overwound and bent. Nucleic Acids Res. 1989;17(7):2541–56.

Fedeles BI, Singh V, Delaney JC, Li D, Essigmann JM. The AlkB family of Fe(II)/alpha-ketoglutarate-dependent dioxygenases: repairing nucleic acid alkylation damage and beyond. J Biol Chem. 2015;290(34):20734–42. doi:10.1074/jbc.R115.656462.

Fleissner E, Borek E. A new enzyme of RNA synthesis: RNA methylase. Proc Natl Acad Sci U S A. 1962;48:1199–203.

Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods. 2010;7(6):461–5. doi:10.1038/nmeth.1459.

Fu Y, Jia G, Pang X, Wang RN, Wang X, Li CJ, et al. FTO-mediated formation of N6-hydroxymethyladenosine and N6-formyladenosine in mammalian RNA. Nat Commun. 2013;4:1798. doi:10.1038/ncomms2822.

Fu Y, Luo GZ, Chen K, Deng X, Yu M, Han D, et al. N(6)-methyldeoxyadenosine marks active transcription start sites in chlamydomonas. Cell. 2015;161(4):879–92. doi:10.1016/j.cell.2015.04.010.

Fujikawa N, Kurumizaka H, Nureki O, Tanaka Y, Yamazoe M, Hiraga S, et al. Structural and biochemical analyses of hemimethylated DNA binding by the SeqA protein. Nucleic Acids Res. 2004;32(1):82–92. doi:10.1093/nar/gkh173.

Gama-Sosa MA, Midgett RM, Slagel VA, Githens S, Kuo KC, Gehrke CW, et al. Tissue-specific differences in DNA methylation in various mammals. Biochim Biophys Acta. 1983;740(2):212–9.

Geier GE, Modrich P. Recognition sequence of the dam methylase of Escherichia coli K12 and mode of cleavage of Dpn I endonuclease. J Biol Chem. 1979;254(4):1408–13.

Gerken T, Girard CA, Tung YC, Webby CJ, Saudek V, Hewitson KS, et al. The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. Science (New York, NY). 2007;318(5855):1469–72. doi:10.1126/science.1151710.

Glickman BW. Spontaneous mutagenesis in Escherichia coli strains lacking 6-methyladenine residues in their DNA: an altered mutational spectrum in dam- mutants. Mutat Res. 1979; 61(2):153–62.

Glickman B, van den Elsen P, Radman M. Induced mutagenesis in dam- mutants of Escherichia coli: a role for 6-methyladenine residues in mutation avoidance. Mol Gen Genet. 1978;163(3):307–12.

Goedecke K, Pignot M, Goody RS, Scheidig AJ, Weinhold E. Structure of the N6-adenine DNA methyltransferase M.TaqI in complex with DNA and a cofactor analog. Nat Struct Biol. 2001;8(2):121–5. doi:10.1038/84104.

Gold M, Hurwitz J. The enzymatic methylation of ribonucleic acid and deoxyribonucleic acid. V. Purification and properties of the deoxyribonucleic acid-methylating activity of Escherichia Coli. J Biol Chem. 1964;239:3858–65.

Gold M, Hurwitz J, Anders M. The enzymatic methylation of RNA and DNA. Biochem Biophys Res Commun. 1963;11:107–14.

Gommers-Ampt JH, Borst P. Hypermodified bases in DNA. FASEB J. 1995;9(11):1034–42.

Gommers-Ampt JH, Van Leeuwen F, de Beer AL, Vliegenthart JF, Dizdaroglu M, Kowalak JA, et al. beta-D-glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan T. brucei. Cell. 1993;75(6):1129–36.

Gorovsky MA, Hattman S, Pleger GL. (6 N)methyl adenine in the nuclear DNA of a eucaryote, Tetrahymena pyriformis. J Cell Biol. 1973;56(3):697–701.

Graham MW, Larkin PJ. Adenine methylation at dam sites increases transient gene expression in plant cells. Transgenic Res. 1995;4(5):324–31.

Greer EL, Beese-Sims SE, Brookes E, Spadafora R, Zhu Y, Rothbart SB, et al. A histone methylation network regulates transgenerational epigenetic memory in C. elegans. Cell Rep. 2014;7(1):113–26. doi:10.1016/j.celrep.2014.02.044.

Greer EL, Becker B, Latza C, Antebi A, Shi Y. Mutation of C. elegans demethylase spr-5 extends transgenerational longevity. Cell Res. 2016;26(2):229–38. doi:10.1038/cr.2015.148.

Greer EL, Blanco MA, Gu L, Sendinc E, Liu J, Aristizabal-Corrales D, et al. DNA Methylation on N(6)-Adenine in C. elegans. Cell. 2015b;161(4):868–78. doi:10.1016/j.cell.2015.04.005.

Grosjean H. Nucleic acids are not boring long polymers of only four types of nucleotides: a guided tour. In: Grosjean H, editor. DNA and RNA modi cation enzymes: structure, mechanism, function and evolution. New York: CRC Press; 2009. p. 1–18.

Grosjean H. RNA modification: the Golden Period 1995–2015. RNA. 2015;21(4):625–6. doi:10.1261/rna.049866.115.

Guarne A, Zhao Q, Ghirlando R, Yang W. Insights into negative modulation of E. coli replication initiation from the structure of SeqA-hemimethylated DNA complex. Nat Struct Biol. 2002;9(11):839–43. doi:10.1038/nsb857.

Hagerman KR, Hagerman PJ. Helix rigidity of DNA: the meroduplex as an experimental paradigm. J Mol Biol. 1996;260(2):207–23. doi:10.1006/jmbi.1996.0393.

Hassel M, Cornelius MG, Vom Brocke J, Schmeiser HH. Total nucleotide analysis of Hydra DNA and RNA by MEKC with LIF detection and 32P-postlabeling. Electrophoresis. 2010;31(2):299–302. doi:10.1002/elps.200900458.

Hattman S, Kenny C, Berger L, Pratt K. Comparative study of DNA methylation in three unicellular eucaryotes. J Bacteriol. 1978;135(3):1156–7.

Heindell HC, Liu A, Paddock GV, Studnicka GM, Salser WA. The primary sequence of rabbit alpha-globin mRNA. Cell. 1978;15(1):43–54.

Hongay CF, Orr-Weaver TL. Drosophila Inducer of MEiosis 4 (IME4) is required for Notch signaling during oogenesis. Proc Natl Acad Sci U S A. 2011;108(36):14855–60. doi:10.1073/pnas.1111577108.

Horton JR, Liebert K, Hattman S, Jeltsch A, Cheng X. Transition from nonspecific to specific DNA interactions along the substrate-recognition pathway of dam methyltransferase. Cell. 2005;121(3):349–61. doi:10.1016/j.cell.2005.02.021.

Horton JR, Liebert K, Bekes M, Jeltsch A, Cheng X. Structure and substrate recognition of the Escherichia coli DNA adenine methyltransferase. J Mol Biol. 2006;358(2):559–70. doi:10.1016/j.jmb.2006.02.028.

Hotchkiss RD. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. J Biol Chem. 1948;175(1):315–32.

Huang W, Xiong J, Yang Y, Liu SM, Yuan BF, Feng YQ. Determination of DNA adenine methylation in genomes of mammals and plants by liquid chromatography/mass spectrometry. Royal Soc Chem Adv. 2015;5:64046–54.

Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature. 2010;466(7310):1129–33. doi:10.1038/nature09303.

Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science (New York, NY). 2011;333(6047):1300–3. doi:10.1126/science.1210597.

Iyer LM, Abhiman S, Aravind L. Natural history of eukaryotic DNA methylation systems. Prog Mol Biol Transl Sci. 2011;101:25–104. doi:10.1016/B978-0-12-387685-0.00002-0.

Jabbari K, Caccio S, Pais de Barros JP, Desgres J, Bernardi G. Evolutionary changes in CpG and methylation levels in the genome of vertebrates. Gene. 1997;205(1–2):109–18.

Jackson JP, Lindroth AM, Cao X, Jacobsen SE. Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature. 2002;416(6880):556–60. doi:10.1038/nature731.

Janulaitis A, Klimasauskas S, Petrusyte M, Butkus V. Cytosine modification in DNA by BcnI methylase yields N4-methylcytosine. FEBS Lett. 1983;161(1):131–4.

Jia G, Fu Y, Zhao X, Dai Q, Zheng G, Yang Y, et al. N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat Chem Biol. 2011;7(12):885–7. doi:10.1038/nchembio.687.

Johnson TB, Coghill RD. The discovery of 5-methyl-cytosine in tuberculinic acid, the nucleic acid of the Tubercle bacillus. J Am Chem Soc. 1925;47:2838–44.

Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, Callis J, et al. The SRA methyl-cytosine-binding domain links DNA and histone methylation. Curr Biol. 2007;17(4):379–84. doi:10.1016/j.cub.2007.01.009.

Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012;13(7):484–92. doi:10.1038/nrg3230.

Kakutani T, Munakata K, Richards EJ, Hirochika H. Meiotically and mitotically stable inheritance of DNA hypomethylation induced by ddm1 mutation of Arabidopsis thaliana. Genetics. 1999;151(2):831–8.

Kamat SS, Fan H, Sauder JM, Burley SK, Shoichet BK, Sali A, et al. Enzymatic deamination of the epigenetic base N-6-methyladenine. J Am Chem Soc. 2011;133(7):2080–3. doi:10.1021/ja110157u.

Karrer KM, VanNuland TA. Methylation of adenine in the nuclear DNA of Tetrahymena is internucleosomal and independent of histone H1. Nucleic Acids Res. 2002;30(6):1364–70.

Katz DJ, Edwards TM, Reinke V, Kelly WG. A C. elegans LSD1 demethylase contributes to germline immortality by reprogramming epigenetic memory. Cell. 2009;137(2):308–20.

Kornberg A, Zimmerman SB, Kornberg SR, Josse J. Enzymatic synthesis of deoxyribonucleic acid. Influence of bacteriophage T2 on the synthetic pathway in host cells. Proc Natl Acad Sci U S A. 1959;45(6):772–85.

Kornberg SR, Zimmerman SB, Kornberg A. Glucosylation of deoxyribonucleic acid by enzymes from bacteriophage-infected Escherichia coli. J Biol Chem. 1961;236:1487–93.

Kozdon JB, Melfi MD, Luong K, Clark TA, Boitano M, Wang S, et al. Global methylation state at base-pair resolution of the Caulobacter genome throughout the cell cycle. Proc Natl Acad Sci U S A. 2013;110(48):E4658–67. doi:10.1073/pnas.1319315110.

Koziol MJ, Bradshaw CR, Allen GE, Costa AS, Frezza C, Gurdon JB. Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. Nat Struct Mol Biol. 2016;23(1):24–30. doi:10.1038/nsmb.3145.

Krais AM, Cornelius MG, Schmeiser HH. Genomic N(6)-methyladenine determination by MEKC with LIF. Electrophoresis. 2010;31(21):3548–51. doi:10.1002/elps.201000357.

Lahue RS, Su SS, Modrich P. Requirement for d(GATC) sequences in Escherichia coli mutHLS mismatch correction. Proc Natl Acad Sci U S A. 1987;84(6):1482–6.

Letunic I, Bork P. Interactive tree of life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res. 2011;39(Web Server issue):W475–8. doi:10.1093/nar/gkr201.

Lichtsteiner S, Schibler U. A glycosylated liver-specific transcription factor stimulates transcription of the albumin gene. Cell. 1989;57(7):1179–87.

Lindahl T, Nyberg B. Heat-induced deamination of cytosine residues in deoxyribonucleic acid. Biochemistry. 1974;13(16):3405–10.

Linn S, Arber W. Host specificity of DNA produced by Escherichia coli, X. In vitro restriction of phage fd replicative form. Proc Natl Acad Sci U S A. 1968;59(4):1300–6.

Liu J, Yue Y, Han D, Wang X, Fu Y, Zhang L, et al. A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat Chem Biol. 2014;10(2):93–5. doi:10.1038/nchembio.1432.

Low DA, Weyand NJ, Mahan MJ. Roles of DNA adenine methylation in regulating bacterial gene expression and virulence. Infect Immun. 2001;69(12):7197–204. doi:10.1128/IAI.69.12.7197-7204.2001.

Lu M, Campbell JL, Boye E, Kleckner N. SeqA: a negative modulator of replication initiation in E. coli. Cell. 1994;77(3):413–26.

Luo GZ, Blanco MA, Greer EL, He C, Shi Y. DNA N-methyladenine: a new epigenetic mark in eukaryotes? Nat Rev. 2015;16(12):705–10. doi:10.1038/nrm4076.

Luria SE, Human ML. A nonhereditary, host-induced variation of bacterial viruses. J Bacteriol. 1952;64(4):557–69.

Lyko F, Ramsahoye BH, Jaenisch R. DNA methylation in Drosophila melanogaster. Nature. 2000;408(6812):538–40. doi:10.1038/35046205.

Machnicka MA, Milanowska K, Osman Oglou O, Purta E, Kurkowska M, Olchowik A, et al. MODOMICS: a database of RNA modification pathways--2013 update. Nucleic Acids Res. 2013;41(Database issue):D262–7. doi:10.1093/nar/gks1007.

Macon JB, Wolfenden R. 1-Methyladenosine. Dimroth rearrangement and reversible reduction. Biochemistry. 1968;7(10):3453–8.

Malagnac F, Bartee L, Bender J. An Arabidopsis SET domain protein required for maintenance but not establishment of DNA methylation. EMBO J. 2002;21(24):6842–52.

Malygin EG, Lindstrom Jr WM, Schlagman SL, Hattman S, Reich NO. Pre-steady state kinetics of bacteriophage T4 dam DNA-[N(6)-adenine] methyltransferase: interaction with native (GATC) or modified sites. Nucleic Acids Res. 2000;28(21):4207–11.

Marinus MG. Adenine methylation of Okazaki fragments in Escherichia coli. J Bacteriol. 1976;128(3):853–4.

Marinus MG, Lobner-Olesen A. DNA methylation. Ecosal Plus. 2014;6(1). doi:10.1128/ecosalplus.ESP-0003-2013.

Marinus MG, Morris NR. Isolation of deoxyribonucleic acid methylase mutants of Escherichia coli K-12. J Bacteriol. 1973;114(3):1143–50.

Marinus MG, Morris NR. Biological function for 6-methyladenine residues in the DNA of Escherichia coli K12. J Mol Biol. 1974;85(2):309–22.

Martin C, Zhang Y. Mechanisms of epigenetic inheritance. Curr Opin Cell Biol. 2007;19(3):266–72. S0955-0674(07)00054-3 [pii]. doi:10.1016/j.ceb.2007.04.002.

Mason SF. Purine Studies. Part II. The Ultra-violet absorption spectra of some mono- and poly-substituted purines. J Chem Soc. 1954:2071–81.

McClelland M. Selection against dam methylation sites in the genomes of DNA of enterobacteriophages. J Mol Evol. 1984;21(4):317–22.

Meselson M, Yuan R. DNA restriction enzyme from E. coli. Nature. 1968;217(5134):1110–4.

Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. Cell. 2012;149(7):1635–46. doi:10.1016/j.cell.2012.05.003.

Mills JB, Hagerman PJ. Origin of the intrinsic rigidity of DNA. Nucleic Acids Res. 2004;32(13):4055–9. doi:10.1093/nar/gkh740.

Montero LM, Filipski J, Gil P, Capel J, Martinez-Zapater JM, Salinas J. The distribution of 5-methylcytosine in the nuclear genome of plants. Nucleic Acids Res. 1992;20(12):3207–10.

Murchie AI, Lilley DM. Base methylation and local DNA helix stability. Effect on the kinetics of cruciform extrusion. J Mol Biol. 1989;205(3):593–602.

Murray NE. 2001 Fred Griffith review lecture. Immigration control of DNA in bacteria: self versus non-self. Microbiology. 2002;148(Pt 1):3–20. doi:10.1099/00221287-148-1-3.

Nikolskaya II, Lopatina NG, Chaplygina NM, Debov SS. The host specificity system in Escherichia coli SK. Mol Cell Biochem. 1976;13(2):79–87.

Nikolskaya II, Lopatina NG, Debov SS. On heterogeneity of DNA methylases from Escherichia coli SK cells. Mol Cell Biochem. 1981;35(1):3–10.

Niu Y, Zhao X, Wu YS, Li MM, Wang XJ, Yang YG. N6-methyl-adenosine (m6A) in RNA: an old modification with a novel epigenetic function. Genomics Proteomics Bioinformatics. 2013;11(1):8–17. doi:10.1016/j.gpb.2012.12.002.

Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell. 1999;99(3):247–57.

Parfrey LW, Lahr DJ, Knoll AH, Katz LA. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proc Natl Acad Sci U S A. 2011;108(33):13624–9. doi:10.1073/pnas.1110633108.

Peng S, Padva A, LeBreton PR. Ultraviolet photoelectron studies of biological purines: the valence electronic structure of adenine. Proc Natl Acad Sci U S A. 1976;73(9):2966–8.

Pogolotti Jr AL, Ono A, Subramaniam R, Santi DV. On the mechanism of DNA-adenine methylase. J Biol Chem. 1988;263(16):7461–4.

Pomraning KR, Smith KM, Freitag M. Genome-wide high throughput analysis of DNA methylation in eukaryotes. Methods. 2009;47(3):142–50. doi:10.1016/j.ymeth.2008.09.022.

Posfai G, Szybalski W. A simple method for locating methylated bases in DNA using class-IIS restriction enzymes. Gene. 1988;74(1):179–81.

Pratt K, Hattman S. Nucleosome phasing in Tetrahymena macronuclei. J Protozool. 1983;30(3): 592–8.

Privat E, Sowers LC. Photochemical deamination and demethylation of 5-methylcytosine. Chem Res Toxicol. 1996;9(4):745–50. doi:10.1021/tx950182o.

Proffitt JH, Davie JR, Swinton D, Hattman S. 5-Methylcytosine is not detectable in Saccharomyces cerevisiae DNA. Mol Cell Biol. 1984;4(5):985–8.

Pukkila PJ, Peterson J, Herman G, Modrich P, Meselson M. Effects of high levels of DNA adenine methylation on methyl-directed mismatch repair in Escherichia coli. Genetics. 1983;104(4): 571–82.

Quignard E, Fazakerley GV, Teoule R, Guy A, Guschlbauer W. Consequences of methylation on the amino group of adenine. A proton two-dimensional NMR study of d(GGATATCC) and d(GGm6ATATCC). Eur J Biochem. 1985;152(1):99–105.

Rae PM. Hydroxymethyluracil in eukaryote DNA: a natural feature of the pyrrophyta (dinoflagellates). Science (New York, NY). 1976;194(4269):1062–4.

Rae PM, Spear BB. Macronuclear DNA of the hypotrichous ciliate Oxytricha fallax. Proc Natl Acad Sci U S A. 1978;75(10):4992–6.

Razin A, Razin S. Methylated bases in mycoplasmal DNA. Nucleic Acids Res. 1980;8(6): 1383–90.

Reich NO, Mashhoon N. Kinetic mechanism of the EcoRI DNA methyltransferase. Biochemistry. 1991;30(11):2933–9.

Robbins-Manke JL, Zdraveski ZZ, Marinus M, Essigmann JM. Analysis of global gene expression and double-strand-break formation in DNA adenine methyltransferase- and mismatch repair-deficient Escherichia coli. J Bacteriol. 2005;187(20):7027–37. doi:10.1128/JB.187.20.7027-7037.2005.

Roberts D, Hoopes BC, McClure WR, Kleckner N. IS10 transposition is regulated by DNA adenine methylation. Cell. 1985;43(1):117–30.

Rogers JC, Rogers SW. Comparison of the effects of N6-methyldeoxyadenosine and N5-methyldeoxycytosine on transcription from nuclear gene promoters in barley. Plant J. 1995;7(2):221–33.

Rogers SD, Rogers ME, Saunders G, Holt G. Isolation of mutants sensitive to 2-aminopurine and alkylating agents and evidence for the role of DNA methylation in Penicillium chrysogenum. Curr Genet. 1986;10(7):557–60.

Romanov GA, Vanyushin BF. Methylation of reiterated sequences in mammalian DNAs. Effects of the tissue type, age, malignancy and hormonal induction. Biochim Biophys Acta. 1981;653(2):204–18.
Russell DW, Hirata RK. The detection of extremely rare DNA modifications. Methylation in dam- and hsd- Escherichia coli strains. J Biol Chem. 1989;264(18):10787–94.
Saparbaev M, Laval J. Excision of hypoxanthine from DNA containing dIMP residues by the Escherichia coli, yeast, rat, and human alkylpurine DNA glycosylases. Proc Natl Acad Sci U S A. 1994;91(13):5873–7.
Sarnacki SH, Castaneda Mdel R, Noto Llana M, Giacomodonato MN, Valvano MA, Cerquetti MC. Dam methylation participates in the regulation of PmrA/PmrB and RcsC/RcsD/RcsB two component regulatory systems in Salmonella enterica serovar Enteritidis. PLoS One. 2013;8(2):e56474. doi:10.1371/journal.pone.0056474.
Sater MR, Lamelas A, Wang G, Clark TA, Roltgen K, Mane S, et al. DNA methylation assessed by SMRT sequencing is linked to mutations in Neisseria meningitidis isolates. PLoS One. 2015;10(12):e0144612. doi:10.1371/journal.pone.0144612.
Sedgwick B, Bates PA, Paik J, Jacobs SC, Lindahl T. Repair of alkylated DNA: recent advances. DNA Repair (Amst). 2007;6(4):429–42. doi:10.1016/j.dnarep.2006.10.005.
Seisenberger S, Andrews S, Krueger F, Arand J, Walter J, Santos F, et al. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol Cell. 2012;48(6):849–62. doi:10.1016/j.molcel.2012.11.001.
Shapiro R, Klein RS. The deamination of cytidine and cytosine by acidic buffer solutions. Mutagenic implications. Biochemistry. 1966;5(7):2358–62.
Shen L, Song CX, He C, Zhang Y. Mechanism and function of oxidative reversal of DNA and RNA methylation. Annu Rev Biochem. 2014;83:585–614. doi:10.1146/annurev-biochem-060713-035513.
Slater S, Wold S, Lu M, Boye E, Skarstad K, Kleckner N. E. coli SeqA protein binds oriC in two different methyl-modulated reactions appropriate to its roles in DNA replication initiation and origin sequestration. Cell. 1995;82(6):927–36.
Smith JD, Arber W, Kuhnlein U. Host specificity of DNA produced by Escherichia coli. XIV. The role of nucleotide methylation in in vivo B-specific modification. J Mol Biol. 1972;63(1):1–8.
Srivastava R, Gopinathan KP, Ramakrishnan T. Deoxyribonucleic acid methylation in mycobacteria. J Bacteriol. 1981;148(2):716–9.
Stein R, Gruenbaum Y, Pollack Y, Razin A, Cedar H. Clonal inheritance of the pattern of DNA methylation in mouse cells. Proc Natl Acad Sci U S A. 1982;79(1):61–5.
Stephens C, Reisenauer A, Wright R, Shapiro L. A cell cycle-regulated bacterial DNA methyltransferase is essential for viability. Proc Natl Acad Sci U S A. 1996;93(3):1210–4.
Sternglanz H, Bugg CE. Conformation of N6-methyladenine, a base involved in DNA modification: restriction processes. Science (New York, NY). 1973;182(4114):833–4.
Su SS, Modrich P. Escherichia coli mutS-encoded protein binds to mismatched DNA base pairs. Proc Natl Acad Sci U S A. 1986;83(14):5057–61.
Sugimoto K, Takeda S, Hirochika H. Transcriptional activation mediated by binding of a plant GATA-type zinc finger protein AGP1 to the AG-motif (AGATCCAA) of the wound-inducible Myb gene NtMyb2. Plant J. 2003;36(4):550–64.
Sun Q, Huang S, Wang X, Zhu Y, Chen Z, Chen D. N(6) -methyladenine functions as a potential epigenetic mark in eukaryotes. Bioessays. 2015;37(11):1155–62. doi:10.1002/bies.201500076.
Sundheim O, Talstad VA, Vagbo CB, Slupphaug G, Krokan HE. AlkB demethylases flip out in different ways. DNA Repair (Amst). 2008;7(11):1916–23. doi:10.1016/j.dnarep.2008.07.015.
Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science (New York, NY). 2009;324(5929):930–5. doi:10.1126/science.1170116.
Tawa R, Ueno S, Yamamoto K, Yamamoto Y, Sagisaka K, Katakura R, et al. Methylated cytosine level in human liver DNA does not decline in aging process. Mech Ageing Dev. 1992;62(3):255–61.

Theil EC, Zamenhof S. Studies on 6-methylaminopurine (6-methyladenine) in bacterial deoxyribonucleic acid. J Biol Chem. 1963;238:3058–64.

Trewick SC, Henshaw TF, Hausinger RP, Lindahl T, Sedgwick B. Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. Nature. 2002;419(6903):174–8. doi:10.1038/nature00908.

Tronche F, Rollier A, Bach I, Weiss MC, Yaniv M. The rat albumin promoter: cooperation with upstream elements is required when binding of APF/HNF1 to the proximal element is partially impaired by mutation or bacterial methylation. Mol Cell Biol. 1989;9(11):4759–66.

Tsuchiya H, Matsuda T, Harashima H, Kamiya H. Cytokine induction by a bacterial DNA-specific modified base. Biochem Biophys Res Commun. 2005;326(4):777–81. doi:10.1016/j.bbrc.2004.11.115.

Unger G, Venner H. Remarks on minor bases in spermatic desoxyribonucleic acid. Hoppe Seylers Z Physiol Chem. 1966;344(4):280–3.

Urbonavicius J, Skouloubris S, Myllykallio H, Grosjean H. Identification of a novel gene encoding a flavin-dependent tRNA:m5U methyltransferase in bacteria--evolutionary implications. Nucleic Acids Res. 2005;33(13):3955–64. doi:10.1093/nar/gki703.

Urig S, Gowher H, Hermann A, Beck C, Fatemi M, Humeny A, et al. The Escherichia coli dam DNA methyltransferase modifies DNA in a highly processive reaction. J Mol Biol. 2002;319(5):1085–96. doi:10.1016/S0022-2836(02)00371-6.

van den Born E, Omelchenko MV, Bekkelund A, Leihne V, Koonin EV, Dolja VV, et al. Viral AlkB proteins repair RNA damage by oxidative demethylation. Nucleic Acids Res. 2008;36(17):5451–61. doi:10.1093/nar/gkn519.

Van Etten JL, Schuster AM, Girton L, Burbank DE, Swinton D, Hattman S. DNA methylation of viruses infecting a eukaryotic Chlorella-like green alga. Nucleic Acids Res. 1985;13(10):3471–8.

Vanyushin BF, Belozersky AN, Kokurina NA, Kadirova DX. 5-methylcytosine and 6-methylamino-purine in bacterial DNA. Nature. 1968;218(5146):1066–7.

Vanyushin BF, Tkacheva SG, Belozersky AN. Rare bases in animal DNA. Nature. 1970; 225(5236):948–9.

von Freiesleben U, Rasmussen KV, Schaechter M. SeqA limits DnaA activity in replication from oriC in Escherichia coli. Mol Microbiol. 1994;14(4):763–72.

Wagner I, Capesius I. Determination of 5-methylcytosine from plant DNA by high-performance liquid chromatography. Biochim Biophys Acta. 1981;654(1):52–6.

Wallecha A, Munster V, Correnti J, Chan T, van der Woude M. Dam- and OxyR-dependent phase variation of agn43: essential elements and evidence for a new role of DNA methylation. J Bacteriol. 2002;184(12):3338–47.

Wang X, Lu Z, Gomez A, Hon GC, Yue Y, Han D, et al. N6-methyladenosine-dependent regulation of messenger RNA stability. Nature. 2014;505(7481):117–20. doi:10.1038/nature12730.

Wang X, Zhao BS, Roundtree IA, Lu Z, Han D, Ma H, et al. N(6)-methyladenosine modulates messenger RNA translation efficiency. Cell. 2015;161(6):1388–99. doi:10.1016/j.cell.2015.05.014.

Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet. 2005;37(8):853–62. doi:10.1038/ng1598.

Wei YF, Carter KC, Wang RP, Shell BK. Molecular cloning and functional analysis of a human cDNA encoding an Escherichia coli AlkB homolog, a protein involved in DNA alkylation damage repair. Nucleic Acids Res. 1996;24(5):931–7.

Willis DB, Granoff A. Frog virus 3 DNA is heavily methylated at CpG sequences. Virology. 1980;107(1):250–7.

Wion D, Casadesus J. N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. Nat Rev Microbiol. 2006;4(3):183–92. doi:10.1038/nrmicro1350.

Wold S, Boye E, Slater S, Kleckner N, Skarstad K. Effects of purified SeqA protein on oriC-dependent DNA replication in vitro. EMBO J. 1998;17(14):4158–65. doi:10.1093/emboj/17.14.4158.

Wu JC, Santi DV. Kinetic and catalytic mechanism of HhaI methyltransferase. J Biol Chem. 1987;262(10):4778–86.

Wu TP, Wang T, Seetin MG, Lai Y, Zhu S, Lin K, et al. DNA methylation on N-adenine in mammalian embryonic stem cells. Nature. 2016. doi:10.1038/nature17640.

Wyatt GR. Occurrence of 5-methylcytosine in nucleic acids. Nature. 1950;166(4214):237–8.

Wyatt GR, Cohen SS. A new pyrimidine base from bacteriophage nucleic acids. Nature. 1952;170(4338):1072–3.

Yamaki H, Ohtsubo E, Nagai K, Maeda Y. The oriC unwinding by dam methylation in Escherichia coli. Nucleic Acids Res. 1988;16(11):5067–73.

Yang CG, Yi C, Duguid EM, Sullivan CT, Jian X, Rice PA, et al. Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. Nature. 2008;452(7190):961–5. doi:10.1038/nature06889.

Yoder JA, Soman NS, Verdine GL, Bestor TH. DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe. J Mol Biol. 1997;270(3):385–95. doi:10.1006/jmbi.1997.1125.

Yue Y, Liu J, He C. RNA N6-methyladenosine methylation in post-transcriptional gene expression regulation. Genes Dev. 2015;29(13):1343–55. doi:10.1101/gad.262766.115.

Yuki H, Kawasaki H, Imayuki A, Yajima T. Determination of 6-methyladenine in DNA by high-performance liquid chromatography. J Chromatogr. 1979;168(2):489–94.

Zaleski P, Wojciechowski M, Piekarowicz A. The role of Dam methylation in phase variation of Haemophilus influenzae genes involved in defence against phage infection. Microbiology. 2005;151(Pt 10):3361–9. doi:10.1099/mic.0.28184-0.

Zelinkova E, Paulicek M, Zelinka J. Modification methylase M. Sau3239I from Streptomyces aureofaciens 3239. FEBS Lett. 1990;271(1–2):147–8.

Zhang G, Huang H, Liu D, Cheng Y, Liu X, Zhang W, et al. N(6)-methyladenine DNA modification in Drosophila. Cell. 2015;161(4):893–906. doi:10.1016/j.cell.2015.04.018.

Zheng G, Dahl JA, Niu Y, Fedorcsak P, Huang CM, Li CJ, et al. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol Cell. 2013;49(1):18–29. doi:10.1016/j.molcel.2012.10.015.

Zhou J, Wan J, Gao X, Zhang X, Jaffrey SR, Qian SB. Dynamic m(6)A mRNA methylation directs translational control of heat shock response. Nature. 2015;526(7574):591–4. doi:10.1038/nature15377.

Pathways of DNA Demethylation

Wendy Dean

Abstract

The regulation of the genome relies on the epigenome to instruct, define and restrict the activities of growth and development. Among the cohort of epigenetic instructions, DNA methylation is perhaps the best understood. In most mammals, cycles of the addition and removal of DNA methylation constitute phases of reprogramming when the developing embryo must negotiate lineage defining and developmental commitment events. In these instances, the DNA methylation instruction is often removed, thereby allowing a change in permission for future development and a return to a more plastic and pluripotent state. Because of this, the germ line, upon demethylation, can give rise to gametes that are fully functional across generations and poised for totipotency. This return to a less differentiated state can also be achieved experimentally. The loss of DNA methylation constitutes one of the significant barriers to induced pluripotency and is a prerequisite for the generation of iPS cells. Taking fully differentiated cells, such as skin cells, and turning back the developmental clock heralded a technological breakthrough discovery in 2006 (Takahashi and Yamanaka 2006) with unprecedented promise in regenerative medicine. In this chapter, the mechanistic possibilities for DNA demethylation will be described in the context of natural and experimentally induced epigenetic reprogramming. The balance of the maintenance of this heritable mark together with its timely removal is essential for lifelong health and may be a key in our understanding of ageing.

W. Dean
Epigenetics Programme, The Babraham Institute, Cambridge CB22 3AT, UK
e-mail: wendy.dean@babraham.ac.uk

© Springer International Publishing Switzerland 2016

A. Jeltsch, R.Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*,
Advances in Experimental Medicine and Biology 945,
DOI 10.1007/978-3-319-43624-1_11

Abbreviations

5caC	5-Carboxylcytosine
5fC	5-Formylcytosine
5hmC	5-Hydroxymethylcytosine
5mC	5-Methylcytosine
A	Adenosine
AID	Activation-induced deaminase
AICDA	Activation-induced cytosine deaminase
Ape1	Apurinic/apyrimidinic (AP) endonuclease 1
APOBEC	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like
BER	Base excision repair
C	Cytosine
CpA	Cytosine–adenosine dinucleotide
CpG	Cytosine–guanosine dinucleotide
CGI	CpG islands
CXXC	Zinc finger protein-binding domain to non-methylated CpG
CHH	Asymmetric DNA methylation
DNA	Deoxyribonucleic acid
ES cells	Embryonic stem cells
DNMTs	DNA methyltransferases, Dnmt1, DNA (cytosine-5)-methyltransferase 1; Dnmt1o, DNA (cytosine-5)-methyltransferase 1 oocyte form; Dnmt1s, DNA (cytosine-5)-methyltransferase 1 somatic form; Dnmt3a, DNA (cytosine-5)-methyltransferase 3a
Dnmt3b	DNA (cytosine-5)-methyltransferase 3b
Dnmt3L	DNA (cytosine-5)-methyltransferase 3-like, E6.5, embryonic day 6.5
E13.5	Embryonic day thirteen
EGFP	Enhanced green fluorescent protein
ELP1	Elongator complex protein 1
ELP3	Elongator complex protein 3
ELP4	Elongator complex protein 4
GV	Germinal vesicle
GVOs	Germinal vesicle oocytes
GSE	Gonad-specific expression
G	Guanosine
IAP	Intracisternal A particles, IF, immunofluorescence
iPS cells	Induced pluripotent stem cells
H3K9me2	Histone H3 lysine 9 dimethylation
KO	Knockout, MBD2, methyl-CpG-binding domain 2
MBD4	Methyl-CpG-binding domain 4
NGS	Next-generation sequencing
NER	Nucleotide excision repair
Np95	Nuclear protein 95
RFTD	Replication foci targeting domain
PARP1	Poly-ADP-ribose polymerase 1
PRC2	Polycomb repressive complex

PGCs	Primordial germ cells
RRBS	Reduced representational bisulphite sequencing
RNA	Ribonucleic acid
RNAi	RNA interference
SAM	S-Adenosyl-L-methionine
siRNA	Small interfering RNA
SMUG1	Single-strand selective monofunctional uracil DNA glycosylase 1
SNT	Somatic nuclear transfer
T	Thymine
TDG	Thymine DNA glycosylase
TET1–3	Ten-eleven translocation 1, 2 or 3
U	Uracil
UNG2, ZGA	Zygotic genome activation

1 DNA Methylation: One Building Block of the Epigenome

If the four bases of DNA constitute the building blocks of life, then the variation between cell types arising from a common origin, as is the case during the cellular differentiation of the zygote, requires another layer to interpret the underlying genetic code. This is the role of the epigenome, and one of the critical components of this is the modification of cytosine by the addition of a methyl group in the C5 position of the cytosine base. Early TLC analysis of total hydrolysates of calf thymus DNA identified a satellite to the cytosine spot on chromatograms, and this was dubbed the fifth base of DNA (Hotchkiss 1948). Methylcytosine was found frequently but differed in its abundance according to tissue and species (Kothari and Shankar 1976; Ehrlich et al. 1982). Further evaluation placed this methylation in a 5mCpG dinucleotide context (Grippo et al. 1968; Russell et al. 1976) which has dominated our thinking relative to the genomic nature of DNA methylation and the mechanism of its modulation, both gain and loss, over the last 40 years.

The machinery which adds DNA methylation to cytosine has been extensively studied and is largely comprised of a small group of enzymatic activities, the DNA methyltransferases (Dnmts), as well as non-enzymatic family members and cofactors including S-adenosyl-L-methionine as the methyl group donor. These details are investigated and outlined in detail elsewhere in this book. Moreover DNA methyltransferases possess important multiple functional partnerships via the replication foci targeting domain (RFTD) and CXXC domain of Dnmt1 and the PHD and PWWP domains of the Dnmt3 enzymes, which bind chromatin and other proteins. Together with a wide selection of splice variants, especially among Dnmt3b, these activities possess an expanding potential to create a wide division of labour for DNA methylation in the vertebrate genome. However, these mechanisms and activities are also at the centre of how DNA methylation may be lost.

In any discussion of DNA methylation reprogramming, it is important to consider the role of DNA methylation. DNA methylation is regarded to have two non-exclusive functions. Very early on, it was associated with the induction of changes

in the state of cellular differentiation in experimental situations where 5-azacytidine, an inhibitor of DNA methylation, resulted in the generation of muscle cells from non-muscle precursors (Constantinides et al. 1977). This highlighted two epigenetic principles, which dominated the field for many years: that transcription was negatively correlated with DNA methylation and that the loss of DNA methylation resulted in the loss of heritable cell fate (Jaenisch and Bird 2003). The gene regulatory role has led to the widespread concept that DNA methylation silences gene expression. In a developing embryo, this can then be tied to the explanation for the sequential methylation of genes no longer required as cellular differentiation and lineage commitment take place. In a highly related fashion, the second role for DNA methylation is in genome defence where the vast non-genic content of retrotransposons is kept silenced by the presence of DNA methylation (Bestor 1998). In addition, DNA methylation is highly relevant to the silencing of families of repeat sequences including the centromeric satellite repeats essential for chromosome integrity and segregation in mitosis (Lehnertz et al. 2003; Saksouk et al. 2014). These differential requirements add to the complexity of reprogramming and serve to explain the multifaceted mechanisms, both passive and active, that may be required during genomic resetting as part of reprogramming.

The link between DNA demethylation and development is longstanding. Model systems of development, such as embryonal carcinoma cells (a lineage-specific cell line isolated from teratocarcinomas that exhibits developmental plasticity), male germ cells (sperm) (which are highly methylated) as a starting point of development and yolk sac as a more developmentally advancement tissue, established that lineages undergo DNA demethylation during development (Razin et al. 1984). A loss of methylation had already been reported to be present in transformed cells and in tumours (Riggs and Jones 1983). In the simplest of terms, DNA demethylation can be consigned to two clearly defined categories, active demethylation that required a 'demethylase' and passive demethylation that assumed that there was ongoing DNA replication, but the DNA methylation maintenance machinery was not operational. These very clear distinctions formed the basis for understanding all examples of demethylation, at times creating impassioned schisms among the chief proponents.

In mammals, DNA methylation is reprogrammed in two significant phases: one that takes place during the establishment of the germ line and the second immediately post-fertilisation as part of an enduring cycling that perpetuates genetic material across generations (Reik et al. 2001). In this respect, the idea of the immortality of the germ line is formed, driven by the reiterative cycling of DNA methylation and demethylation. As such, much of the investigation of passive and active demethylation has focused around these cycles especially in the mouse.

As mentioned, DNA demethylation is critically required in the germ line where, in mammals, the imprinted genes must be reset. These genes are expressed in a parent-of-origin manner and hence must undergo erasure of the DNA methylation that marks the imprinted regulatory regions in each generation (Reik et al. 2001). Coupled to the erasure of imprinting, a number of germ line-specific genes, such as *Dazl*, must undergo methylation erasure in order to be expressed, thus ensuring the complete development of gametes (Seisenberger et al. 2012). However, the kinetics of the erasure is tightly regulated, such that methylation from germ line-specific

genes is erased earlier (E10.5) than that of imprinted genes (from E11.5) (Maatouk et al. 2006). These multiple trajectories may even require both active and passive mechanisms to operate simultaneously across the genome as well as at specific loci.

A second period of demethylation is initiated immediately at fertilisation. While quantitative and qualitative differences have been reported across Mammalia, the mechanisms are largely similar (Dean et al. 2001; Young and Beaujean 2004). Owing to the extensive resource tools available to investigate mechanistic pathways, the mouse model has been most comprehensively studied and has revealed a number of activities involved in passive and active demethylation. In recent years, technological advancements have been key in revealing a much richer and more intricate temporal change in DNA methylation during the preimplantation period of development. In this window, the fully mature gametes give way to the newly formed zygote, coupling their remodelling with the loss of methylation from both the male and female genomes and setting up a hypomethylated landscape to coincide with the first lineage decisions that delimit the preimplantation stage of development. The early observation that the paternal contribution to the zygote, the remodelled sperm, and thereafter the male pronucleus underwent a rapid and extensive loss of methylation prior to replication inspired the field that an activity, a demethylase, might be present (Oswald et al. 2000; Mayer et al. 2000; Santos et al. 2002).

2 Active Demethylation: The Hunt for the 'Demethylase'

The definition of active demethylation is straightforward. It refers to the loss of DNA methylation in the absence of replication. In dividing cells, this requires the proof that this loss must occur outside of 'S phase' or that it is not related to the machinery required for DNA replication per se, if it occurs in S phase. In order to fulfil this criterion, the mechanism must conform to something where it can be envisioned that the methyl group is directly 'snipped' away from the base leaving behind cytosine. Another variation might include stripping away of the base and leaving behind an abasic site, which then needs some type of DNA repair to restore the cytosine. This mechanism is exemplified by the plant family of activities, DEMETER and ROSI, DNA glycosylases, which recognise the methylated base and remove it via the base excision repair (BER) pathway. Enzymes which bind DNA and catalyse chemical reactions on it do so by distorting, bending or kinking the B-form helices. Some of these enzymes, such as the DNA methyltransferases, achieve this by flipping the base out and modifying it, thus creating a methylated nucleotide. This leads to the idea that the reverse reaction might occur in order for demethylation to take place. Base flipping is also thought to occur in some of the glycosylases in order to gain access to the DNA bases (Cheng and Roberts 2001). These ideas supplied a variety of possibilities and gave clues to finding and confirming DNA demethylases in animals.

The study of DNA methylation and its modulation has been ongoing for more than 40 years. Early studies using chromatography and nearest neighbour analysis could only have reported on total changes in 5-methylcytosine (5mC) (Bird 1980).

This was superseded by the use of methylation-sensitive restriction endonucleases at CpG dinucleotides, followed by Southern blot that gave context assuming that the genes or regions had been cloned (Cedar et al. 1979). Breakthroughs came in the combination of differential methylation sensitivity of restriction enzymes and the polymerase chain reaction (Herman et al. 1995) together with the base pair resolution afforded by bisulphite sequencing (BS-seq) (Frommer et al. 1992; Clark et al. 1994). Coupled with the refinements in sample sizes by orders of magnitude, the field experienced a true renaissance with the mouse genome project in 2003 and with the advent of next-generation sequencing.

3 Direct Demethylation

If DNA methylation is negatively correlated with gene expression, then the identification of a demethylase that regulates expression would provide an integral part of the mechanistic understanding of gene regulation, and this enzyme might serve important roles therapeutically in the targeted re-expression of key genes involved in developmental and cellular processes. Among the most obvious candidate genes for such application are the tumour suppressor genes invariably silenced by inappropriate acquisition of DNA methylation found in precancerous and transformed cells of tumours (Herman et al. 1995). In nearly all tumours, cells have a signature where tumour suppressor genes are hypermethylated and repressed or silent while the remainder of the genome is hypomethylated (Gama-Sosa et al. 1983; Feinberg and Vogelstein 1983). In this context, a concerted effort to identify and characterise a demethylase had attracted much interest beyond other biological roles establishing the methylation landscape during development.

The publication of the discovery of such a demethylation activity in 1999 was received with great interest and a degree of scepticism in the biomedical field and the burgeoning field of epigenetics (Bhattacharya et al. 1999). The existence of such an activity was highly controversial, as it had been regarded that the direct loss of a methyl group from cytosine was thermodynamically unfavourable. This new protein possessed the hallmark of such an activity together with a methyl-CpG-binding domain, and demethylation of methylcytosine was observed in the CpG context from artificially methylated plasmid DNA. Moreover, in the course of this enzymatic reaction, the methyl group was removed from the five position of the cytosine base and released as methanol, a stable leaving group. This methyl-binding protein had been cloned previously and identified as MBD2. Despite the flurry of interest in this molecule, independent verification of direct demethylation catalysed by MBD2, and hence demethylase activity of this protein, has not been reported (Ng et al. 1999; Wade et al. 1999; Boeke et al. 2000). Aside from this controversy, MBD2 has been implicated in demethylation processes in a wide variety of autoimmune diseases in mice and humans. Whether MBD2 functions as a demethylase during the immediate post-fertilisation period prior to the first zygotic S phase seems unlikely, as oocytes null for MBD2 undergo loss of DNA methylation from the male pronucleus in a kinetically similar manner to that of the control population (Santos et al. 2002).

4 Indirect Loss of DNA Methylation

4.1 Deamination

One of the mechanisms that have shaped the mammalian genome is the process whereby the amine group (−NH_2) of the nucleotide base is lost by deamination. This is particularly relevant for cytosine (C) and especially the case for its modified version 5mC, as its deamination leads to the natural base thymidine leading to a substitution known as the C-to-T transition. The propensity of this alteration has had a significant impact on the mammalian genome and its unique composition (Bird 1980). Among these genomic signatures is the suppression of the expected frequency of CpG dinucleotides. The altered frequency and the genomic arrangement of CpG have significant implications for the regulation of DNA methylation. The genomic frequency of CpG methylation is closer to 1 in 40 rather than the predicted 1 in 16. This altered frequency occurs as a consequence, over evolutionary time, of the transition mutation of methylcytosine to thymidine owing to the hydrolytic deamination of the cytosine base where the fidelity of the repair process is apt to slip. In addition, this transition forms the basis of the chemical modification used in bisulphite mutagenesis, where 5mC is read as T (Frommer et al. 1992). These two concepts led to the suggestion that endogenous activity, either enzymatic or chemical, such as through hydrolytic deamination, could serve as candidates capable of achieving functional loss of DNA methylation.

Spontaneous hydrolytic deamination of cytosine occurs ordinarily at a rate of eight bases per cell per hour, corresponding to 100 bases per cell per day (Alberts et al. 2002). Whether this mechanism is competent to clear the DNA methylation from the male pronucleus of the fertilised oocyte during the active period of demethylation is not known. However, the rate of this process seems far too slow to account for the kinetics of demethylation observed in immunofluorescence (IF) and by contemporary molecular analysis. In contrast, enzymatic deamination coupled to DNA repair processes offers a diverse and more feasible solution to explain the rate and magnitude of active loss of methylation in the zygote. In mammals, a small family of highly related activities achieves deamination of cytosine and methylcytosine. Cytosine deaminases are encoded by three enzymatic families, (1) APOBEC1, APOBEC2 and APOBEC3, of which APOBEC2 is non-enzymatic (Conticello et al. 2005), and (2) the activation-induced cytosine deaminase (AID), which takes its name from its functional role in somatic hypermutation and class switch recombination that permits the repertoire expansion of antibody diversification upon B-cell activation as part of the adaptive immune response (Longerich et al. 2006). The APOBEC family of enzymes are broadly involved in RNA editing and as such serve a role in genome defence via engagement to restrict the movement of non-LTR and LTR retrotransposable elements, including long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs) (Chiu and Greene 2008). In the case of retroviral integrations via transposons and nuclear APOBEC3, this establishes the idea that an early evolutionary coupling of the expansion of the genome together with activities that may alter and defend the functional integrity of this self-same genome. This highlights some of the dynamic facets of the genome and the epigenome as they could have evolved in parallel.

4.2 Oxidation Mechanisms

In the same way as deamination modifies 5mC to thymidine, the oxidation of the 5-methyl group can rapidly convert 5mC into 5-hydroxymethylcytosine (5hmC) without any further alteration to the double-stranded DNA molecule. Under favourable conditions, this process may proceed through further iterative oxidative steps to form 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Ito et al. 2010; Ko et al. 2010). Should a decarboxylase be present or simply a repair cascade be triggered, this pathway could mediate the conversion of 5mC to cytosine and the demethylation will be complete. The oxidative portion of this process is accomplished variously by the ten-eleven-translocation (TET) family, comprised of three active forms of the TET1–3 enzymes, which all share a conserved function using 5mC as a substrate in conjunction with 5-α-ketoglutarate and Fe^{2+} as cofactors (Kriaucionis and Heintz 2009; Tahiliani et al. 2009). As such, these enzymes are influenced by the metabolic state of the cell, and the concentration of each oxidative intermediate is dependent on a number of other cofactors including vitamins A and C, as well as the oxygen concentration of the tissue or cells (Blaschke et al. 2013; Chen et al. 2013). TET proteins may operate together with other 'loss of methylation' pathways creating a much more complex and diversified means of epigenetic regulation. Interestingly, these enzymes may serve wider biological roles beyond their canonical enzymatic function in keeping with their evolutionary secondment into a demethylation pathway.

Site-specific, feature-specific or ultimately genome-wide loss of DNA methylation may require more than a single pathway to achieve a change in the epigenome. While some activities operate exclusively during the replicative portion of the cell cycle, others may act in a redundant fashion through a combination of DNA modifications, leading to loss of methylation often by repair-coupled processes. This is the case with cytosine deamination and oxidative modification of 5-methylcytosine (Fig. 1). The order of modifications may also change, permitting yet further complexity and specificity at target loci. This may widen and even accelerate the rate of repair or alteration with additional functional consequences. For example, deamination of 5mC leads to T, which may be further altered or repaired, e.g. by thymidine DNA glycosylase (TDG). Alternatively, 5mC may be oxidised to 5hmC, which may be deaminated to give 5-hydroxymethyluracil (5hmU). In either case, the loss of methylation would be the outcome upon resolution of the repair process.

5 Chromatin Remodelling, DNA Replication and Repair: The Epigenetic Triumvirate

Whenever chromatin is remodelling, there is a likelihood of the occurrence of endogenous DNA damage. As such, the genome-wide remodelling during epigenetic reprogramming is particularly susceptible to DNA damage, especially in replicating cells where the replication fork can become a powerful nucleation site for activities focused on inheritance of genetic and epigenetic information. Both natural and experimental reprogramming, which activates new transcriptional requirements,

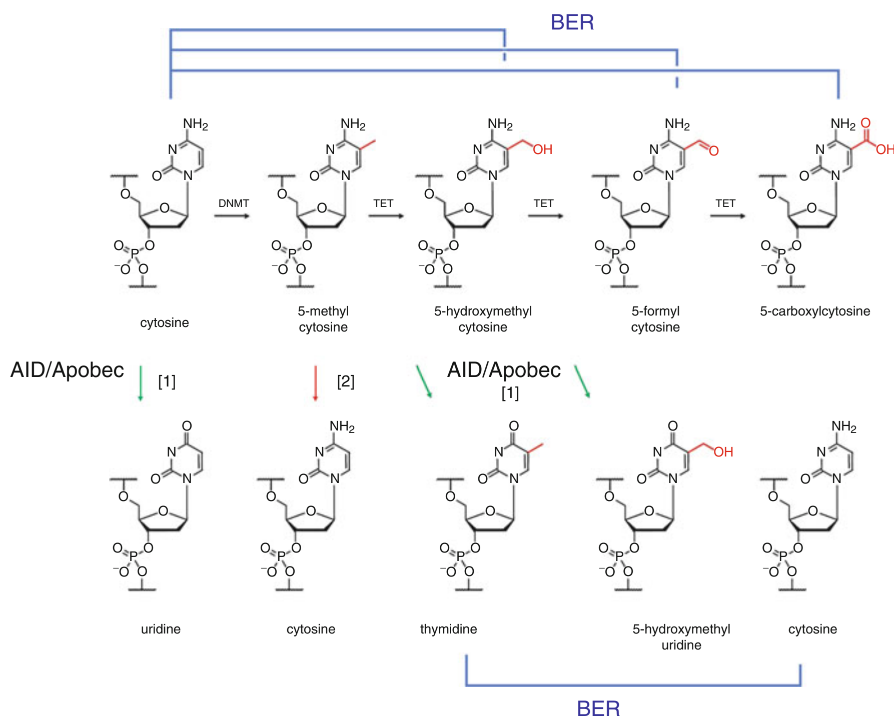


Fig. 1 Mechanisms of active demethylation. Cytosine is methylated by DNA methyltransferases (Dnmts) as an integral part in the development of mammals. These processes are also accompanied by demethylation via a number of potential pathways. Cytosine may be altered by [*1*] AID/APOBEC-mediated deamination, leaving uracil, and 5-methylcytosine may likewise undergo deamination to thymidine. This pathway may be continued to restore the cytosine residue by one of a number of base excision repair (BER) pathways. The 5mC residue may also be directly demethylated to cytosine, although this is regarded as enzymatically unfavourable [*2*]. Another mechanism to facilitate loss of methylation occurs via TET-mediated oxidative alteration of the methyl group to give 5-hydroxymethylcytosine (5hmC) that may undergo further iterative oxidation to 5-formylcytosine (5fC). In the absence of glycosylases that can detect this mismatch, 5fC may be further oxidised to 5-carboxylcytosine (5caC). BER mechanisms may aid in the repair of the cytosine base at each step of the process. The loss of methylation from cytosine may involve both pathways. In this case, 5mC is first oxidised to 5hmC and subsequently deaminated to 5-hydroxymethyluracil

may trigger the need for a DNA damage response. Collectively, these may result in functional demethylation and could become a source of cellular heterogeneity in development.

In the window defined in the immediate post-fertilisation period in the mouse, male and female gametes are remodelled in order to form a functional metaphase plate, the first of all the subsequent cellular divisions for life in mammals. This remodelling is essentially required in order to re-equilibrate the chromatin of the somatic-like oocyte with that of the protamine-configured sperm nucleus (Brewer et al. 1999). In order to package the paternal component, the sperm, into a functional structure able to fertilise an ovulated oocyte, the genome must be confined to

the dehydrated toroidal structure configured by protamines, something more akin to a virus than a cell (Braun 2001). This is the state of play when the largest cell, the oocyte, meets the smallest, the highly methylated sperm. Very shortly following fusion and sperm penetration, remodelling of this genome ensues in the shared environment of the oocyte. Indeed, within 1 h of fertilisation, the protamine-encased chromatin has been replaced, and the sperm becomes configured with oocyte-derived histone proteins (McLay and Clarke 2003). In this early remodelling phase, significant DNA methylation is lost by paternal alleles (Dean et al. 2001; Santos et al. 2002). In contrast, the female pronucleus that forms on completion of meiosis is less subject to loss of DNA methylation (Santos et al. 2002); quantification of the female compartment even under the most stringent conditions indicates a wide variance in DNA methylation and suggests that chromatin turnover also takes place on maternal alleles, but there is no net change (Santos et al. 2013).

Despite the formation of a histone-based chromatin in the male pronucleus, the male and female components of the zygote remain epigenetically different (Santos et al. 2005; Probst et al. 2007). Post-translational modifications (PTMs) of the chromatin retain differences between the pronuclei beyond that of DNA methylation. The PTMs leave the female with most of the modifications found in a somatic cell, while notably the male pronucleus has neither H3K9me2 nor H3K9me3 (Santos et al. 2005; van der Heijden et al. 2005). Both of these marks are critical due to the coupling of DNA methylation to these modifications, usually associated with transcriptional silencing, and they are essential for chromosome stability (Guenatri et al. 2004). The absence of these marks from the chromatin in the male pronucleus facilitates the loss of DNA methylation. In turn, this suggested that one or both of these marks may well be associated with protecting the maternal genome from extensive DNA demethylation (Nakamura et al. 2007; Szabo and Pfeifer 2012). Interestingly, the male pronucleus accumulates 5hmC only once the cell cycle progresses to S phase with only newly acquired 5mC converted to 5hmC, while the female pronucleus has acquired significantly less of the 5hmC modification during oogenesis (Santos et al. 2013; Amouroux et al. 2016). This rapid paternal loss of methylation was attributed to the presence of TET3 owing to its overt abundance in oocytes (Iqbal et al. 2011). However, whether this 5mC turnover is required for subsequent development and what protects 5mC from demethylation in the female pronucleus are not fully known. Several groups have knocked out the TET3 catalytic activity with varying phenotypic results (Gu et al. 2011; Peat et al. 2014). All agree that the abrogation of the catalytic activity results in the loss of 5hmC as measured by IF signal in the male pronucleus, but a concomitant increase in 5mC was not always observed (Wossidlo et al. 2011; Santos et al. 2013). The absolute loss of catalytic TET3 is non-viable and homozygotes die shortly after birth. Maternal-specific oocyte deletion reveals a more subtle nonetheless variable effect. Several reports highlight that the failure to generate 5hmC per se is not responsible for the developmental defects, e.g. neonatal sub-lethality due to TET3 haploinsufficiency in maternally deficient oocytes (Inoue et al. 2015; Tsukada et al. 2015). Interestingly, genome-wide DNA methylation analysis of the TET3 knockout revealed an unexpected role for this protein in protecting CpG islands from DNA methylation (Peat et al. 2014). This intriguing result reminds us that the machinery

of demethylation is complex and multidimensional. The protection of the female pronucleus may give us an insight into the overall mechanism of the regulation of the loss of DNA methylation and its role.

DNA repair has long been implicated in the process of active demethylation especially in the mouse zygote. The mammalian oocyte is replete with many of the activities that are required across the myriad repair pathways supplying roles in maintaining the genome during replication and remodelling (Zheng et al. 2005; Derijck et al. 2006, 2008). In the special case of active CpG demethylation, more focus has been placed on activities that can read and repair the mismatched derivatives of CG base pairs generated by modification of the cytosine.

Several groups have followed avenues of investigation implicating BER or NER pathways in active demethylation using both gain- and loss-of-function approaches (Wossidlo et al. 2010; Hajkova et al. 2010). Inhibition of critical activities in BER, such as Parp1 [poly(ADP-ribose) polymerase family] and Ape1 (apurinic/apyrimidinic endonuclease), was able to attenuate DNA demethylation with significantly higher levels of DNA methylation in the paternal pronucleus, as judged by 5mC staining (Hajkova et al. 2010). Staining of XRCC1 (X-ray repair complementing defective repair in Chinese hamster cells 1), a protein known to bind to ssDNA breaks, revealed high levels of bound protein exclusive to the male pronucleus. The ssDNA breaks were detectable from an early pronuclear stage (PN3) on, which coincides with the onset of DNA demethylation. Collectively, these pathways account for some of the early demethylation in the zygote; yet these studies fail to address the question of what initiates the pathways. Dnmt3a and Dnmt3b are abundant in oocytes and early embryos (Hirasawa and Sasaki 2009; Lucifero et al. 2007) and have been implicated in deamination roles in human cells (Kangaspeska et al. 2008; Metivier et al. 2008). However, whether they serve as the upstream initiator of the cascade has not been tested directly. Recent reports highlight that de novo methylation is required for the replication-dependent generation of 5hmC in the male pronucleus, reinforcing its functional status in the zygote as an independent mark (Amouroux et al. 2016). In contrast, there may not be an activity required to initiate this demethylation; instead conformational changes of the sperm chromatin structure during chromatin remodelling in the zygote may be able to trigger DNA damage response mechanisms leading to loss of DNA methylation (McLay and Clarke 2003).

Apart from deamination- and oxidation-linked processes of DNA demethylation, the transcription elongator complex protein 3 (ELP3) and the family members ELP1 and ELP4 have been identified in candidate-based screens for demethylating activities and have been reported to act via an otherwise unknown mechanism to effect active demethylation. Using live cell imaging as a platform to capture the event, the authors expressed a construct (CXXC–GFP) able to act as a readout for the state of DNA methylation in the zygote. The CXXC–GFP fusion protein reporter is able to bind to unmethylated CpG-rich regions and hence can follow the kinetics of DNA methylation over time. On injection of the CXXC–GFP reporter, signal accumulated in the male pronucleus. When an RNAi to ELP was co-injected, no signal was observed in the male pronucleus, suggesting an ELP3-dependent loss of methylation at this time (Okada et al. 2010). However, the mechanism of this paternal-specific demethylation is not immediately obvious. Moreover, in light of

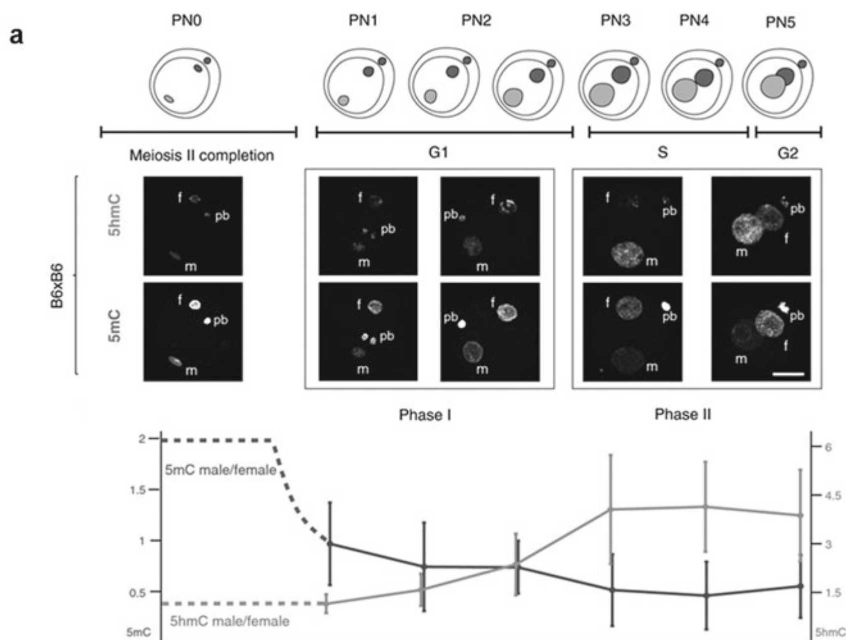
reports suggesting that some turnover in DNA methylation occurs in the female compartment, the exclusive accumulation of the reporter CXXC–GFP in the male might not be expected (Guo et al. 2014). Nonetheless, given that this phase of development is characterised by transcriptional silencing (Bouniol-Baly et al. 1999), the suggestion that ELPs play a role is somewhat counterintuitive. A functional connection may lie in that these factors share a radical S-adenosyl-methionine (SAM) domain and hence are potentially able to modify the target C5 of cytosine, when in the presence of [3Fe–4S] + clusters, with the generation of a neutral leaving group, like formaldehyde (Broderick et al. 2014; Wu and Zhang 2014). This idea is reminiscent of the mechanism proposed for the activity of MBD2 where, in the absence of any cofactor, nucleophilic attack by H_2O was proposed to lead to demethylation and the production of methanol (Bhattacharya et al. 1999).

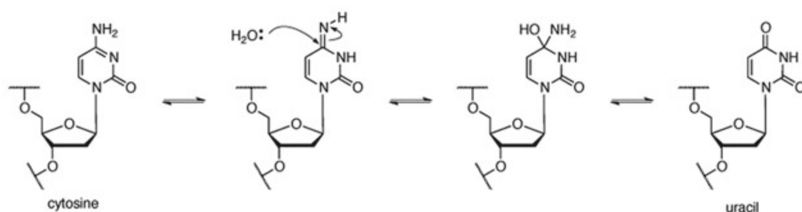
Several groups have reported the kinetics of the loss of 5mC and tied this to the acquisition of 5hmC but have failed to notice that there were clearly two phases (Wossidlo et al. 2011; Santos et al. 2013). In the first phase, paternal-specific active demethylation proceeds in the absence of DNA replication and as such without any accumulation of 5hmC. In this window, remodelling of the paternal pronucleus includes the introduction of replication-independent histone variants. The second phase is characterised by the onset of replication and the concomitant accumulation of 5hmC exclusively in the paternal compartment.

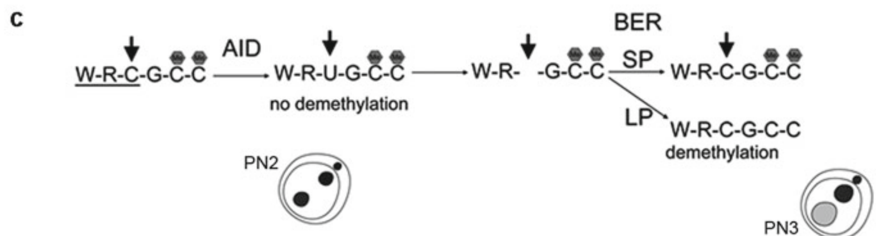
Santos et al. (2013) used a genetic approach to test the possibility that both AID and TET3 could act in the 'demethylation' pathways (Santos et al. 2013). Investigating the long overlooked two phases of demethylation, the dynamics and magnitude of the loss of DNA methylation was measured by semi-quantitative IF methods first in the wild-type (WT) zygotes (Fig. 2a) and then using a constitutive AID-null mutant. In the null mutant, the early phase of demethylation proceeded as in the WT, but by the post-replication stage, PN5 residual gain of methylation was evident. This suggested that an AID-dependent loss of DNA methylation from the male pronucleus was possible in the first cell cycle in the mouse. But how can this temporal shift in the kinetics be explained? The answer came by way of an ingenious mechanism

Fig. 2 Active demethylation during reprogramming in the zygote. Following fertilisation, the maternal and paternal pronuclei are remodelled along two independent pathways. (**a**) Chromatin is reconfigured in stages across time point defined by the size and position of the respective pronuclei (PN_0 to PN_5). The paternal pronucleus is observed to lose methylation in two distinct phases. In the first phase, 5mC is rapidly lost with little or no change in 5hmC marking the end of G_1. The onset of phase II at S phase of the cell cycle (PN_3–PN_5) coincides with the appearance of 5hmC, a process that is TET3 dependent. Across both phases, the maternal pronucleus retains methylation and possesses very low levels of 5hmC, and little overall change takes place. The maternal pronucleus resists loss of methylation in a Stella-/PGC7-dependent manner. One mechanism proposed to achieve loss of methylation in the paternal pronucleus is mediated by deamination depicted in (**b**). Here cytosine is modified through the loss of the NH_2 group to uracil, triggering a BER cascade involving a short-patch repair (SP) (**c**). The AID target sequence WRC marks the position of the deaminated cytosine in a place adjacent to 5mC residues. It is envisioned that in phase I of demethylation, the cytosine base is deaminated, and the resection of the local 5mC only takes place in phase II in a replication-activated process resulting in functional demethylation. Schematic depictions of fertilised oocytes indicate that initially both compartments have similar methylation levels that lead to paternal-specific loss of 5mC following short patch repair via BER

that fits well the temporal licensing of AID and the biochemical substrate data for AID's catalytic preference. Aid ordinarily works during G1 of the cell cycle, a time consistent with an active demethylation process. Moreover, although AID can work on 5mC, its preferred substrate is C as the enzymatic turnover does not work







very efficiently when the C5 position is derivatised especially by oxidation to hmC (Nabel et al. 2012). Thus, it was envisioned that deamination of C to U (Fig. 2b), in close proximity to 5mC, would undergo repair via a long-patch BER mechanism and would in turn result in functional demethylation (Fig. 2c). This mechanism would involve uracil DNA glycosylases (UNG and SMUG) rather than TDG, a result confirmed using a maternally deleted TDG null zygote (Santos et al. 2013). UNG and SMUG are abundant in oocytes while TDG is not readily detectable. UNG-null zygotes showed a reduced demethylation by IF (Santos et al. 2013) and molecular analysis (Xue et al. 2016).

This study also followed closely the kinetics of the specific acquisition of 5hmC as a function of the reciprocal loss of 5mC from the male pronucleus. Conditional deletion of TET3 from the oocyte, together with experiments inhibiting DNA replication with aphidicolin, clearly demonstrated that 5hmC was only present from S phase on and that it was coincident with, but not coupled to, DNA replication (Santos et al. 2013). Together, this groundbreaking study asserted that multiple and independent pathways were involved in paternal demethylation in the zygote. These findings have been conclusively and independently confirmed in a recent study using genetic and biochemical approaches together with IF where both 5mC and 5hmC were measured using mass spectrometry (Amouroux et al. 2016; Okamoto et al. 2016). In support of these suggested mechanisms, two additional studies have also proposed that long-patch repair may facilitate regional loss of 5mC (Franchini et al. 2014) when adjacent to an AID-binding site, i.e. WRC (W = A/T; R = A/G) or via a similar mechanism through repair of 5-hydroxymethyluracil (hmU) arising from oxidative modification of T by TET1 (Pfaffeneder et al. 2014). One of the 11 glycosylases in human cells, Nei endonuclease VIII-like 1 (NEIL1) and NEIL3, has DNA glycosylase/lyase activity towards mismatched uracil and thymine, in particular in U:C and T:C mismatches, and it specifically binds 5hmC, suggesting that it acts as a specific reader of 5hmC in advance of repair. As such, activities targeting both C and T via separate pathways may result in regional demethylation.

During the late foetal stages and early on in postnatal stages, oogenesis proceeds, and DNA methylation is reacquired in mature germ cells. This re-establishes the essential DNA methylation at imprinted loci needed for growth and development during the post-implantation phase. Some of this methylation is modified becoming 5hmC in a TET1-dependent manner. This establishes, in part, the chromatin configuration for maternal alleles at fertilisation. Together with H3K9me2, the maternal chromatin is actively protected from post-fertilisation demethylation by the presence of the protein Stella/PGC7 that forms a resistant chromatin configuration that is maintained post-fertilisation (Nakamura et al. 2012). While Stella is found in both male and female pronuclei, the binding of the protein in chromatin confers the protective facet against demethylation (Nakamura et al. 2007). Maternal deletion of Stella in the oocyte reinstates a substrate that is subject to TET3-dependent demethylation in both compartments and leads to lethality early on in development. Thus, demethylation can be affected by the ability to physically bind and modify the substrate and may not simply be a function of presence or absence of the demethylating activity.

Interestingly, there is little evidence that the loss of methylation from paternal alleles is required. Paternal-specific demethylation has been widely observed in mammals often to varying degrees, yet an absolute demonstration of its requirement remains elusive. Experimental reconstitution of the zygote using round spermatids, to represent the paternal component, does not undergo remodelling, neither is demethylation of the paternal pronucleus observed, but still can give rise to a full-term mouse (Kishigami et al. 2006; Polanski et al. 2008). Indeed, this raises the issue of whether the extensive remodelling and dynamic loss of DNA methylation in the paternal compartment is rather a secondary consequence of remodelling. The fertilisation of oocytes from species of mammals that do not ordinarily undergo overt demethylation of the paternal pronucleus does so when they are artificially fertilised with sperm from other species. Thus, in sheep oocytes, where there is little or no paternal-specific active demethylation, mouse sperm is extensively remodelled and appears to be demethylated as measured by IF. The reciprocal is also true; the fertilisation of mouse oocytes with sheep sperm reveals extensive remodelling of the sperm nucleus, including the specific loss of methylation (Beaujean et al. 2004; Barnetova et al. 2010).

A new study using highly sensitive mass spectrometry has finally been able to offer an absolute metric for 5mC and 5hmC levels in the zygote thus revealing further complexity. In order to evaluate the potential for changes exclusively from maternal alleles, the authors experimentally generated parthenogenetic embryos. Following active loss from the paternal pronucleus (~40 %) and a small active loss from the female, the overall trajectory for 5mC and 5hmC across preimplantation regresses with a best fit model that follows replication, that is, this is loss of methylation via passive demethylation (Okamoto et al. 2016). This active loss of DNA methylation confirms previous bisulphite analysis that had asserted this process in the maternal compartment as well (Guo et al. 2014). As such, maternal and paternal haploid contributions to the zygote are equalised in genomic 5mC content by 24 h post-fertilisation (Okamoto et al. 2016). Finally, this study found no evidence for the commensurate gain of 5hmC reciprocal to the loss of 5mC claimed in some studies (Wossidlo et al. 2011; Okamoto et al. 2016).

6 Replication-Coupled Loss of DNA Methylation: Passive Demethylation

For the establishment of the germ line, precursor somatic cells must become reprogrammed, such that the somatic gene expression pattern is inhibited giving way to the early expression of the germ cell markers. Somatic cells of the epiblast are highly methylated with approximately 70 % of CpGs marked. While the need for reprogramming of the germ line has been known for several decades, the limitations of the highly restricted cell numbers and the available technology restricted what we could learn about this process (Monk et al. 1987). Early studies were confined to characterising maternally imprinted genes and highly repetitive sequences, including centromeric satellites and repeat family sequences, following their progression

during germ-line maturation in mouse (Hajkova et al. 2002; Lane et al. 2003) and later in other mammalian model systems. Early genome-wide analysis combined histology together with immunofluorescence detection using the 5mC antibody to give a semi-quantitative overview of early germ cells (Seki et al. 2005, 2007). Relying on an Oct4 or Stella/PGC7 – GFP reporters to identify the germ cells – this low-resolution approach was sufficient to establish that at E7.5 the germ cells were highly methylated, but over the next 24-h period, this methylation was significantly reduced and thereafter no longer detectable up to E12.5. While inconclusive regarding a mechanism for the loss of methylation, these studies gave single-cell resolution and hinted at a variable level of methylation likely to arise from a highly repetitive family or families of genomic sequence.

Next-generation sequencing supplied the breakthrough needed for mechanistic understanding. Whole-genome bisulphite sequencing was used to profile the DNA methylation in wild-type mice and in mice carrying a constitutive deletion of AICDA, a protein with a putative role in active demethylation. The scope of this study was limited by the availability of sufficient DNA to generate a library, and hence, E13.5 was elected for the PGC samples. Comparison of the DNA methylation status of WT with AICDA null gave the first high-resolution unbiased epigenome from PGCs (Popp et al. 2010). Interestingly, in the absence of AID, E13.5 female PGCs had three times more methylation (~21 %) than the WT (~7 % CpG methylation). Differences in male PGCs were less dramatic. The effect was genome-wide except for CGI promoters, suggesting that AICDA had some role to play in the reprogramming machinery in germ cells. Despite this breakthrough, little or no resolution of the matter of the nature of the mechanism was forthcoming although AICDA seemed to be involved.

One of the significant limitations to the Popp et al. (2010) study was that the overall genomic coverage was low. However, spurred on by the prospect of an unbiased, base pair resolution method to describe the DNA methylation landscape coupled to solving the issues of small cell number limitations, a comprehensive evaluation of the methylation profile across multiple stages of PGCs was carried out. Oct4-GFP-positive PGCs were isolated from selected stages from E9.5 to E13.5 during PGC development in males and females (Seisenberger et al. 2012). Genome-wide DNA methylation was analysed using BS-seq. High-resolution coverage and bp resolution generated the first unbiased genome-wide survey of DNA methylation in PGCs across development with the potential to answer the question of demethylation via active or passive mechanisms. Combining strand-specific bisulphite sequencing analysis and modelling of stage-specific data, the outcome was clear-cut. Germ-line reprogramming proceeded by a mechanism dominated by passive loss of DNA methylation through progressive replication up to the mitotic/meiotic arrest around E14 (Fig. 3c). These results were confirmed by DNA methylation immunoprecipitation (meDIP) on a microarray (Guibert et al. 2012). This validation was particularly important owing to the ambiguity of calling 5mC and 5hmC by BS-seq. Evaluating the role, if any, of hydroxymethylation became important to resolve the profiles observed during the passive loss of methylation. Primordial germ cells express high levels of TET1 and TET2, and these activities are replication dependent; the 5hmC mark which they generate from the 5mC is not maintained or read at CpG

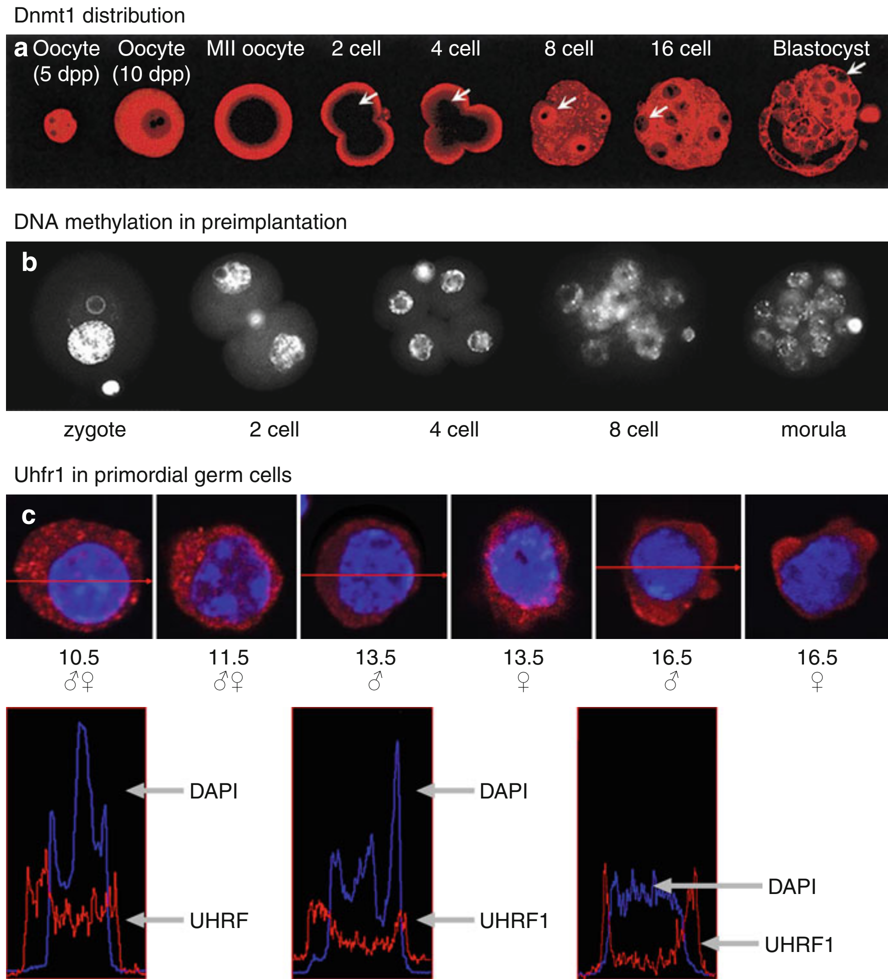


Fig. 3 Passive demethylation during preimplantation and germ-line reprogramming. Very early on, the discovery that Dnmt1 accumulated in the subcortical region of the oocyte and that of preimplantation stage embryos established the expectation that maintenance methylation was interrupted during preimplantation. (**a**) This panel illustrates this principle. Dnmt1 antibody signal is clearly abundant and excluded from the nucleus up until the 8-cell (8C) stage when it is found in the nucleus, thus establishing the conditions for passive demethylation. Beyond the 16-cell stage, the protein is again found excluded from the nucleus, something that also continues up to the blastocyst stage (**b**). Immunofluorescence detection of 5mC suggested that there was a diminishing quantity of DNA methylation over successive replication cycles across preimplantation stages in keeping with passive demethylation. Here the distribution of 5mC is reported from the zygote up to the morula stage. (**c**). A similar mechanism may operate in primordial germ cells. Staining germ cells from E10.5 up to 16.5 indicate that Uhfr1, an obligate binding partner for Dnmt1 and an essential component of the maintenance machinery, is relegated to the cytoplasmic compartment. Despite Dnmt1 residing in the nucleus, the cross-sectional tracing below identifies Uhfr1 (*arrow*) depleted in the nucleus (marked by an arrow; stained with DAPI) and found in the cytoplasm. These conditions are requisite for passive demethylation and in keeping with the genome-wide loss of DNA methylation in the reprogramming of the germ line (*Panel A* is reprinted from Howell et al. (2001) with permission from Elsevier. *Panel B* is taken from Santos et al. (2013) Fig. 1 cited herein)

dinucleotides and hence may passively decline. Loss of 5mC via hydroxylation of the methyl group was thus found to contribute to the erasure of imprinted methylation asynchronously in PGCs between E9.5 and E 10.5, but thereafter the genome average obeyed kinetics of loss of methylation in keeping with passive demethylation (Hackett et al. 2013). In contrast, in vitro generation of differentiated PGCs (iPGCs) where TET1 and TET2 were experimentally depleted or deficient indicated that iPGCs were unaffected by the absence of these activities and the lack of genomic 5hmC (Vincent et al. 2013). However, some loci were found hypermethylated, which necessitated an alternative loss of methylation mechanism to generate functional iPGCs. Indeed, the erasure of methylation in the germ line is sufficiently important that multiple and overlapping redundant mechanisms of erasure have arisen to ensure the process and restrict transgenerational inheritance of epimutations (Hackett et al. 2012). In fact, the extent of the functional redundancy for loss of methylation pathways may even stretch to include multiple and overlapping mechanisms operating even in the same genomic locus (Ohno et al. 2013).

If we can measure the dynamic process of demethylation at high resolution and derive a model that indicates that passive demethylation is the favoured mechanism, can we explain how this mechanism might come about? Kurimoto et al. (2008) isolated Blimp1-positive PGC precursors and profiled RNA from them over the early stages of PGC establishment and commitment in an attempt to understand the basis for passive demethylation. These cells were characterised by the downregulation of pathways associated with DNA methylation maintenance – Dnmt1 and Uhfr1 – and BER pathways (including a role for TDG) in conjunction with the loss of methylation by TET oxidation and AICDA deamination (Kurimoto et al. 2008). The selective staining of staged PGCs for Dnmt1 and Uhfr1 suggested that, while Dnmt1 remained in the nucleus, Uhfr1 was cytoplasmic, a configuration in keeping with passive demethylation (Fig. 3c) (Seisenberger et al. 2012). Whether this exclusion of Uhfr1 despite nuclear Dnmt1 can account for loss of 5mC in PGCs has not been tested yet. However, Dnmt1 is very inefficiently recruited in replicating gonadal germ cells consistent with progressive loss of methylation during maturation (Ohno et al. 2013).

7 Resetting and Erasure of the Germ Line: A Barrier Against Transgenerational Inheritance

7.1 Demethylation During Preimplantation Development

Sanford et al. (1984) showed that sperm was highly methylated, and oocytes were less so and that following fertilisation DNA methylation declined progressively (Sanford et al. 1984). These data illustrate that methylation-sensitive Southern blots, although restricted in the genomic loci which could be interrogated, were nonetheless accurately revealing locus-specific changes during development, which included both hyper- and hypomethylation during this dynamic phase. Monk et al. (1987) confirmed and extended these observations using as few as 200 cells, showing dynamic and temporal-specific regulation of DNA methylation

and demethylation during early mouse development (Monk et al. 1987). However, molecular analysis, at that time, could not supply a genome-wide nuclear architectural perspective. Interestingly, in one of the earliest studies investigating the question of loss of methylation, Howlett and Reik (1991) found very little change in the methylation of L1 repeats between the zygote and blastocyst stages when following maternal chromosomes derived from parthenogenetic mouse embryos (Howlett and Reik 1991).

The first study into the dynamic genome-wide distribution of DNA methylation at the single-cell level was reported not in mouse but in bovine. Using stage-specific chromosome spreads across preimplantation development together with the anti-5mC antibody led to the conclusion that active demethylation was absent and hence that the progressive loss of 5mC occurred by a passive mechanism (Rougier et al. 1998; Bourc'his et al. 2001). The visual resolution of the machinery of maintenance methylation by IF had already predicted such outcome (Fig. 3a). Antibodies raised against Dnmt1 had already identified that the highly abundant protein was largely relegated to the cytoplasmic compartment of the zygote and throughout preimplantation and that it appeared to be tethered to the subcortical region by an active process (Carlson et al. 1992; Cardoso and Leonhardt 1999). This striking and stark distribution offered a mechanism consistent with the observed passive demethylation (Carlson et al. 1992; Howell et al. 2001; Ratnam et al. 2002) up to the blastocyst stage (Fig. 3b).

During preimplantation epigenetic reprogramming, the loss of methylation must be more complex and nuanced in contrast to that of germ-line erasure, because some regions must retain their methylation. This specifically affects germ line differentially methylated regions (DMRs), including those of maternally imprinted genes, which during oogenesis are restored in a Dnmt3a-/Dnmt3L-dependent manner (Kaneda et al. 2010). Transcriptional activity during this period stockpiles transcripts in order to maintain the oocyte's integrity as to function as an immediate store of protein for the completion of meiosis at fertilisation and the initiation of embryogenesis. This includes a number of maternal-specific transcripts that supply the oocyte with an extraordinary amount of protein. Numbered among these proteins are Dnmt1, both the somatic (Dnmt1s) and the oocyte (Dnmt1o) form, and Uhfr1, an essential part of the maintenance machinery.

Advances in molecular profiling of DNA methylation have provided a more detailed understanding of the genome-wide methylation reprogramming during preimplantation in mouse and human. Reduced representation bisulphite sequencing (RRBS) is a technique that offers deeper sequencing of CpG-rich regions with a trade-off of lower genomic coverage. This technique covers 5–10 % of the genome and favours CpG islands and the promoters of genes and hence is ideally suited for methylation profiling of imprinted genes. In contrast to earlier studies done by IF, recent studies in the mouse and human confirmed a rapid and global demethylation of paternal alleles post-fertilisation, but a much more limited loss of methylation at maternal alleles, such that, at the blastocyst stage, some imprinted DMRs were maintained (Smallwood et al. 2011; Guo et al. 2014; Smith et al. 2012). Applying a whole-genome bisulphite sequencing (WGBS) approach revealed further details. Many of the 1600 DMRs inherited from the oocyte were only partially demethylated during

preimplantation indicative of mechanisms conferring resistance to demethylation at play (Kobayashi et al. 2012). A similar study in human blastocysts found that maternal alleles were demethylated to a much lesser extent than in mouse, with notable locus-specific and species-specific differences attributed to both the de novo methylation and demethylation machinery (Okae et al. 2014).

Passive demethylation can occur when the methylation machinery is physically excluded from the substrate or degraded in each replication cycle, but other explanations are possible as well; for example, the loss of methylation may occur when the ability of Dnmts to recognise the substrate is lost (Inoue et al. 2011; Inoue and Zhang 2011). This is the situation that arises in the presence of 5hmC and in the presence of asymmetric methylation of cytosine. Asymmetric DNA methylation has been characterised and mechanistically documented in plants, but it has recently been recognised in mammals, and new biological functions are being discovered especially in non-replicating cells in the brain (Lister et al. 2009). This non-canonical DNA methylation is found in CHH (where H = A, C or T) and CpNpG sequences, which represent a nonsymmetrical and symmetrical sequence context, respectively. The most abundant nonsymmetrical configuration for DNA methylation is CpA, which is found in the male germ line during the foetal maturation period where Dnmt3b is expressed at a time when imprinted methylation is first returned (Ichiyanagi et al. 2011). This mark is temporary, as the mitotic expansion of spermatocytes does not allow for the maintenance of the asymmetrical methylation, and hence it disappears over time. In a similar manner, during the re-establishment of DNA methylation during oogenesis, the abundance of Dnmt3a results in the acquisition of extensive CpA methylation. The absence of replication in oogenesis has a number of important implications. Any accumulation of altered or aberrant bases, ordinarily removed by replication, remains static until after ovulation and fertilisation when the first S phase takes place. As such, the mammalian oocyte may accumulate a significant level of genomic CpA, which remains until after fertilisation when it is progressively diluted away in each replication cycle (Tomizawa et al. 2011; Shirane et al. 2013).

In the zygote and preimplantation embryo, demethylation thus follows a path whereby the 5hmC of the male and the 5CpA in the female are both subject to passive reduction irrespective of the cytoplasmic exclusion of the maintenance machinery, because neither 5hmC nor 5CpA can be maintained on replication (Dean 2014; Okamoto et al. 2016).

8 Experimental Reprogramming and the Loss of DNA Methylation

Changes of the differentiation potential of cells are often associated with quantitative changes in the epigenome, especially in DNA methylation. This is the case in somatic nuclear transfer, where the use of cells with reduced DNA methylation greatly enhances the reprogramming of the somatic nucleus and consequently the cloning efficiency. A similar enhanced efficiency, as a function of reduced DNA methylation, has been observed in the cellular reprogramming involved in the generation of induced pluripotent cells.

Studies to understand the developmental equivalence of embryonic stem cells have been aided by the introduction of small-molecule inhibitors that uncouple the Erk1/Erk2 and the GSK signalling pathway revealing a more in vivo-like ground or 'naive' state now termed '2i' (Ficz et al. 2013). This state differs from the classical or 'primed' state, where cells are maintained in a rich signalling environment in the presence of serum. ESCs can be readily moved between these two states with a remarkably rapid and quantitatively significant change in genomic DNA methylation. This dynamic response in the epigenome's DNA methylation may serve as a useful model to understand developmental reprogramming in the germ line and during preimplantation.

Serum to 2i transitions have been characterised at the RNA and epigenome level of DNA. Genome-wide bisulphite profiling indicated that, while most of the genomic features quickly lost methylation, major satellite repeats of the centromere, the IAP class of retrotransposons and imprinted genes were ostensibly resistant to this change (Ficz et al. 2013). Initial findings from RNA-seq noted a striking downregulation of the de novo methylation machinery, while the maintenance pathway, in particular, Dnmts and Uhfr1, was largely unchanged or upregulated similarly to the oxidative pathway mediated by TETs. Prdm14, an important negative transcriptional regulator of the de novo methyltransferase genes, was profoundly upregulated as well. This upregulation of Prdm14 was regarded as supplying mechanistic understanding for the rapid and widespread demethylation. In this transition, DNA methylation is reduced by 50 % (5mC/C) in 24 h, while 5hmC increases from 0.08 to 0.28 % (5hmC/C) (Yamaji et al. 2013; Leitch et al. 2013; Ficz et al. 2013). However, conditional deletion of both de novo methyltransferases has been tested, but the timeline for reduction of DNA methylation to this extent failed to fit the model. In fact, Dnmt3a and Dnmt3b double null ES cells need an excess of 30 passages to become fully demethylated (Chen et al. 2003). More congruous with the observed rapid decline was the suggestion that this was caused by the erosion in the fidelity of the maintenance process (Ficz et al. 2013). The best fit suggested a combination of all three methods, including TET-mediated oxidative pathways albeit with only a minor effect at most loci.

Interestingly, a recent re-evaluation of the reciprocal loss of methylation and gain of hydroxymethylation in the zygote has uncovered similar principles and advanced our understanding of the complexity which is the reprogramming of DNA methylation in mammals (Amouroux et al. 2016). These integrated global and locus-specific modulations will have significant impact in therapeutic applications for regenerative medicine and in the testing of patient-specific treatment which replies on a modification of the epigenome to restore and correct errors associated with human infertility and in the understanding of nutrition and health in ageing.

References

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. DNA repair. In: Molecular biology of the cell. 4th ed. New York: Garland Science; 2002.

Amouroux R, Nashun B, Shirane K, Nakagawa S, Hill PW, D'Souza Z, Nakayama M, Matsuda M, Turp A, Ndjetehe E, Encheva V, Kudo NR, Koseki H, Sasaki H, Hajkova P. De novo DNA methylation drives 5hmC accumulation in mouse zygotes. Nat Cell Biol. 2016;18(2):225–33. doi:10.1038/ncb3296.

Barnetova I, Fulka H, Fulka Jr J. Epigenetic characteristics of paternal chromatin in interspecies zygotes. J Reprod Dev. 2010;56(6):601–6.

Beaujean N, Taylor JE, McGarry M, Gardner JO, Wilmut I, Loi P, Ptak G, Galli C, Lazzari G, Bird A, Young LE, Meehan RR. The effect of interspecific oocytes on demethylation of sperm DNA. Proc Natl Acad Sci U S A. 2004;101(20):7636–40. doi:10.1073/pnas.0400730101.

Bestor TH. The host defence function of genomic methylation patterns. Novartis Found Symp. 1998;214:187–95; discussion 195–189, 228–132.

Bhattacharya SK, Ramchandani S, Cervoni N, Szyf M. A mammalian protein with specific demethylase activity for mCpG DNA. Nature. 1999;397(6720):579–83. doi:10.1038/17533.

Bird AP. DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res. 1980;8(7):1499–504.

Blaschke K, Ebata KT, Karimi MM, Zepeda-Martinez JA, Goyal P, Mahapatra S, Tam A, Laird DJ, Hirst M, Rao A, Lorincz MC, Ramalho-Santos M. Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. Nature. 2013;500(7461):222–6. doi:10.1038/nature12362.

Boeke J, Ammerpohl O, Kegel S, Moehren U, Renkawitz R. The minimal repression domain of MBD2b overlaps with the methyl-CpG-binding domain and binds directly to Sin3A. J Biol Chem. 2000;275(45):34963–7. doi:10.1074/jbc.M005929200.

Bouniol-Baly C, Hamraoui L, Guibert J, Beaujean N, Szollosi MS, Debey P. Differential transcriptional activity associated with chromatin configuration in fully grown mouse germinal vesicle oocytes. Biol Reprod. 1999;60(3):580–7.

Bourc'his D, Le Bourhis D, Patin D, Niveleau A, Comizzoli P, Renard JP, Viegas-Pequignot E. Delayed and incomplete reprogramming of chromosome methylation patterns in bovine cloned embryos. Curr Biol. 2001;11(19):1542–6.

Braun RE. Packaging paternal chromosomes with protamine. Nat Genet. 2001;28(1):10–2. doi:10.1038/88194.

Brewer LR, Corzett M, Balhorn R. Protamine-induced condensation and decondensation of the same DNA molecule. Science. 1999;286(5437):120–3.

Broderick JB, Duffus BR, Duschene KS, Shepard EM. Radical S-adenosylmethionine enzymes. Chem Rev. 2014;114(8):4229–317. doi:10.1021/cr4004709.

Cardoso MC, Leonhardt H. DNA methyltransferase is actively retained in the cytoplasm during early development. J Cell Biol. 1999;147(1):25–32.

Carlson LL, Page AW, Bestor TH. Properties and localization of DNA methyltransferase in preimplantation mouse embryos: implications for genomic imprinting. Genes Dev. 1992;6(12B):2536–41.

Cedar H, Solage A, Glaser G, Razin A. Direct detection of methylated cytosine in DNA by use of the restriction enzyme MspI. Nucleic Acids Res. 1979;6(6):2125–32.

Cheng X, Roberts RJ, AdoMet-dependent methylation, DNA methyltransferases and base flipping. Nucleic Acids Res. 2001;29:3784–95.

Chen T, Ueda Y, Dodge JE, Wang Z, Li E. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. Mol Cell Biol. 2003;23(16):5594–605.

Chen J, Guo L, Zhang L, Wu H, Yang J, Liu H, Wang X, Hu X, Gu T, Zhou Z, Liu J, Liu J, Wu H, Mao SQ, Mo K, Li Y, Lai K, Qi J, Yao H, Pan G, Xu GL, Pei D. Vitamin C modulates TET1 function during somatic cell reprogramming. Nat Genet. 2013;45(12):1504–9. doi:10.1038/ng.2807.

Chiu YL, Greene WC. The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. Annu Rev Immunol. 2008;26:317–53. doi:10.1146/annurev.immunol.26.021607.090350.

Clark SJ, Harrison J, Paul CL, Frommer M. High sensitivity mapping of methylated cytosines. Nucleic Acids Res. 1994;22(15):2990–7.

Constantinides PG, Jones PA, Gevers W. Functional striated muscle cells from non-myoblast precursors following 5-azacytidine treatment. Nature. 1977;267(5609):364–6.

Conticello SG, Thomas CJ, Petersen-Mahrt SK, Neuberger MS. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. Mol Biol Evol. 2005;22(2):367–77. doi:10.1093/molbev/msi026.

Dean W. DNA methylation and demethylation: a pathway to gametogenesis and development. Mol Reprod Dev. 2014;81(2):113–25. doi:10.1002/mrd.22280.

Dean W, Santos F, Stojkovic M, Zakhartchenko V, Walter J, Wolf E, Reik W. Conservation of methylation reprogramming in mammalian development: aberrant reprogramming in cloned embryos. Proc Natl Acad Sci U S A. 2001;98(24):13734–8. doi:10.1073/pnas.241522698.

Derijck AA, van der Heijden GW, Giele M, Philippens ME, van Bavel CC, de Boer P. gamma-H2AX signalling during sperm chromatin remodelling in the mouse zygote. DNA Repair (Amst). 2006;5(8):959–71. doi:10.1016/j.dnarep.2006.05.043.

Derijck A, van der Heijden G, Giele M, Philippens M, de Boer P. DNA double-strand break repair in parental chromatin of mouse zygotes, the first cell cycle as an origin of de novo mutation. Hum Mol Genet. 2008;17(13):1922–37. doi:10.1093/hmg/ddn090.

Ehrlich M, Gama-Sosa MA, Huang LH, Midgett RM, Kuo KC, McCune RA, Gehrke C. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. Nucleic Acids Res. 1982;10(8):2709–21.

Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature. 1983;301(5895):89–92.

Ficz G, Hore TA, Santos F, Lee HJ, Dean W, Arand J, Krueger F, Oxley D, Paul YL, Walter J, Cook SJ, Andrews S, Branco MR, Reik W. FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. Cell Stem Cell. 2013;13(3):351–9. doi:10.1016/j.stem.2013.06.004.

Franchini DM, Chan CF, Morgan H, Incorvaia E, Rangam G, Dean W, Santos F, Reik W, Petersen-Mahrt SK. Processive DNA demethylation via DNA deaminase-induced lesion resolution. PLoS One. 2014;9(7):e97754. doi:10.1371/journal.pone.0097754.

Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci U S A. 1992;89(5):1827–31.

Gama-Sosa MA, Slagel VA, Trewyn RW, Oxenhandler R, Kuo KC, Gehrke CW, Ehrlich M. The 5-methylcytosine content of DNA from human tumors. Nucleic Acids Res. 1983;11(19):6883–94.

Grippo P, Iaccarino M, Parisi E, Scarano E. Methylation of DNA in developing sea urchin embryos. J Mol Biol. 1968;36(2):195–208.

Gu TP, Guo F, Yang H, Wu HP, Xu GF, Liu W, Xie ZG, Shi L, He X, Jin SG, Iqbal K, Shi YG, Deng Z, Szabo PE, Pfeifer GP, Li J, Xu GL. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature. 2011;477(7366):606–10. doi:10.1038/nature10443.

Guenatri M, Bailly D, Maison C, Almouzni G. Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. J Cell Biol. 2004;166(4):493–505. doi:10.1083/jcb.200403109.

Guibert S, Forne T, Weber M. Global profiling of DNA methylation erasure in mouse primordial germ cells. Genome Res. 2012;22(4):633–41. doi:10.1101/gr.130997.111.

Guo F, Li X, Liang D, Li T, Zhu P, Guo H, Wu X, Wen L, Gu TP, Hu B, Walsh CP, Li J, Tang F, Xu GL. Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. Cell Stem Cell. 2014;15(4):447–58. doi:10.1016/j.stem.2014.08.003.

Hackett JA, Zylicz JJ, Surani MA. Parallel mechanisms of epigenetic reprogramming in the germline. Trends Genet. 2012;28(4):164–74. doi:10.1016/j.tig.2012.01.005.

Hackett JA, Sengupta R, Zylicz JJ, Murakami K, Lee C, Down TA, Surani MA. Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. Science. 2013;339(6118):448–52. doi:10.1126/science.1229277.

Hajkova P, Erhardt S, Lane N, Haaf T, El-Maarri O, Reik W, Walter J, Surani MA. Epigenetic reprogramming in mouse primordial germ cells. Mech Dev. 2002;117(1–2):15–23.

Hajkova P, Jeffries SJ, Lee C, Miller N, Jackson SP, Surani MA. Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. Science. 2010;329(5987):78–82. doi:10.1126/science.1187945.

Herman JG, Merlo A, Mao L, Lapidus RG, Issa JP, Davidson NE, Sidransky D, Baylin SB. Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers. Cancer Res. 1995;55(20):4525–30.

Hirasawa R, Sasaki H. Dynamic transition of Dnmt3b expression in mouse pre- and early post-implantation embryos. Gene Expr Patterns. 2009;9(1):27–30. doi:10.1016/j.gep.2008.09.002.

Hotchkiss RD. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. J Biol Chem. 1948;175(1):315–32.

Howell CY, Bestor TH, Ding F, Latham KE, Mertineit C, Trasler JM, Chaillet JR. Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. Cell. 2001;104(6):829–38.

Howlett SK, Reik W. Methylation levels of maternal and paternal genomes during preimplantation development. Development. 1991;113(1):119–27.

Ichiyanagi K, Li Y, Watanabe T, Ichiyanagi T, Fukuda K, Kitayama J, Yamamoto Y, Kuramochi-Miyagawa S, Nakano T, Yabuta Y, Seki Y, Saitou M, Sasaki H. Locus- and domain-dependent control of DNA methylation at mouse B1 retrotransposons during male germ cell development. Genome Res. 2011;21(12):2058–66. doi:10.1101/gr.123679.111.

Inoue A, Zhang Y. Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. Science. 2011;334(6053):194. doi:10.1126/science.1212483.

Inoue A, Shen L, Dai Q, He C, Zhang Y. Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. Cell Res. 2011;21(12):1670–6. doi:10.1038/cr.2011.189.

Inoue A, Shen L, Matoba S, Zhang Y. Haploinsufficiency, but not defective paternal 5mC oxidation, accounts for the developmental defects of maternal Tet3 knockouts. Cell Rep. 2015;10(4):463–70. doi:10.1016/j.celrep.2014.12.049.

Iqbal K, Jin SG, Pfeifer GP, Szabo PE. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. Proc Natl Acad Sci U S A. 2011;108(9):3642–7. doi:10.1073/pnas.1014033108.

Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature. 2010;466(7310):1129–33. doi:10.1038/nature09303.

Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet. 2003;33(Suppl):245–54. doi:10.1038/ng1089.

Kaneda M, Hirasawa R, Chiba H, Okano M, Li E, Sasaki H. Genetic evidence for Dnmt3a-dependent imprinting during oocyte growth obtained by conditional knockout with Zp3-Cre and complete exclusion of Dnmt3b by chimera formation. Genes Cells. 2010;15(3):169–79. doi:10.1111/j.1365-2443.2009.01374.x.

Kangaspeska S, Stride B, Metivier R, Polycarpou-Schwarz M, Ibberson D, Carmouche RP, Benes V, Gannon F, Reid G. Transient cyclical methylation of promoter DNA. Nature. 2008;452(7183):112–5. doi:10.1038/nature06640.

Kishigami S, Van Thuan N, Hikichi T, Ohta H, Wakayama S, Mizutani E, Wakayama T. Epigenetic abnormalities of the mouse paternal zygotic genome associated with microinsemination of round spermatids. Dev Biol. 2006;289(1):195–205. doi:10.1016/j.ydbio.2005.10.026.

Ko M, Huang Y, Jankowska AM, Pape UJ, Tahiliani M, Bandukwala HS, An J, Lamperti ED, Koh KP, Ganetzky R, Liu XS, Aravind L, Agarwal S, Maciejewski JP, Rao A. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature. 2010;468(7325):839–43. doi:10.1038/nature09586.

Kobayashi H, Sakurai T, Imai M, Takahashi N, Fukuda A, Yayoi O, Sato S, Nakabayashi K, Hata K, Sotomaru Y, Suzuki Y, Kono T. Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. PLoS Genet. 2012;8(1):e1002440. doi:10.1371/journal.pgen.1002440.

Kothari RM, Shankar V. 5-Methylcytosine content in the vertebrate deoxyribonucleic acids: species specificity. J Mol Evol. 1976;7(4):325–9.

Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science. 2009;324(5929):929–30. doi:10.1126/science.1169786.

Kurimoto K, Yabuta Y, Ohinata Y, Shigeta M, Yamanaka K, Saitou M. Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. Genes Dev. 2008;22(12):1617–35. doi:10.1101/gad.1649908.

Lane N, Dean W, Erhardt S, Hajkova P, Surani A, Walter J, Reik W. Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. Genesis. 2003;35(2):88–93. doi:10.1002/gene.10168.

Lehnertz B, Ueda Y, Derijck AA, Braunschweig U, Perez-Burgos L, Kubicek S, Chen T, Li E, Jenuwein T, Peters AH. Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr Biol. 2003;13(14):1192–200.

Leitch HG, McEwen KR, Turp A, Encheva V, Carroll T, Grabole N, Mansfield W, Nashun B, Knezovich JG, Smith A, Surani MA, Hajkova P. Naive pluripotency is associated with global DNA hypomethylation. Nat Struct Mol Biol. 2013;20(3):311–6. doi:10.1038/nsmb.2510.

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009;462(7271):315–22. doi:10.1038/nature08514.

Longerich S, Basu U, Alt F, Storb U. AID in somatic hypermutation and class switch recombination. Curr Opin Immunol. 2006;18(2):164–74. doi:10.1016/j.coi.2006.01.008.

Lucifero D, La Salle S, Bourc'his D, Martel J, Bestor TH, Trasler JM. Coordinate regulation of DNA methyltransferase expression during oogenesis. BMC Dev Biol. 2007;7:36. doi:10.1186/1471-213X-7-36.

Maatouk DM, Kellam LD, Mann MR, Lei H, Li E, Bartolomei MS, Resnick JL. DNA methylation is a primary mechanism for silencing postmigratory primordial germ cell genes in both germ cell and somatic cell lineages. Development. 2006;133(17):3411–8. doi:10.1242/dev.02500.

Mayer W, Niveleau A, Walter J, Fundele R, Haaf T. Demethylation of the zygotic paternal genome. Nature. 2000;403(6769):501–2. doi:10.1038/35000654.

McLay DW, Clarke HJ. Remodelling the paternal chromatin at fertilization in mammals. Reproduction. 2003;125(5):625–33.

Metivier R, Gallais R, Tiffoche C, Le Peron C, Jurkowska RZ, Carmouche RP, Ibberson D, Barath P, Demay F, Reid G, Benes V, Jeltsch A, Gannon F, Salbert G. Cyclical DNA methylation of a transcriptionally active promoter. Nature. 2008;452(7183):45–50. doi:10.1038/nature06544.

Monk M, Boubelik M, Lehnert S. Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. Development. 1987;99(3):371–82.

Nabel CS, Jia H, Ye Y, Shen L, Goldschmidt HL, Stivers JT, Zhang Y, Kohli RM. AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. Nat Chem Biol. 2012;8(9):751–8. doi:10.1038/nchembio.1042.

Nakamura T, Arai Y, Umehara H, Masuhara M, Kimura T, Taniguchi H, Sekimoto T, Ikawa M, Yoneda Y, Okabe M, Tanaka S, Shiota K, Nakano T. PGC7/Stella protects against DNA demethylation in early embryogenesis. Nat Cell Biol. 2007;9(1):64–71. doi:10.1038/ncb1519.

Nakamura T, Liu YJ, Nakashima H, Umehara H, Inoue K, Matoba S, Tachibana M, Ogura A, Shinkai Y, Nakano T. PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. Nature. 2012;486(7403):415–9. doi:10.1038/nature11093.

Ng HH, Zhang Y, Hendrich B, Johnson CA, Turner BM, Erdjument-Bromage H, Tempst P, Reinberg D, Bird A. MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. Nat Genet. 1999;23(1):58–61. doi:10.1038/12659.

Ohno R, Nakayama M, Naruse C, Okashita N, Takano O, Tachibana M, Asano M, Saitou M, Seki Y. A replication-dependent passive mechanism modulates DNA demethylation in mouse primordial germ cells. Development. 2013;140(14):2892–903. doi:10.1242/dev.093229.

Okada Y, Yamagata K, Hong K, Wakayama T, Zhang Y. A role for the elongator complex in zygotic paternal genome demethylation. Nature. 2010;463(7280):554–8. doi:10.1038/nature08732.

Okae H, Chiba H, Hiura H, Hamada H, Sato A, Utsunomiya T, Kikuchi H, Yoshida H, Tanaka A, Suyama M, Arima T. Genome-wide analysis of DNA methylation dynamics during early human development. PLoS Genet. 2014;10(12):e1004868. doi:10.1371/journal.pgen.1004868.

Okamoto Y, Yoshida N, Suzuki T, Shimozawa N, Asami M, Matsuda T, Kojima N, Perry AC, Takada T. DNA methylation dynamics in mouse preimplantation embryos revealed by mass spectrometry. Sci Rep. 2016;6:19134. doi:10.1038/srep19134.

Oswald J, Engemann S, Lane N, Mayer W, Olek A, Fundele R, Dean W, Reik W, Walter J. Active demethylation of the paternal genome in the mouse zygote. Curr Biol. 2000;10(8):475–8.

Peat JR, Dean W, Clark SJ, Krueger F, Smallwood SA, Ficz G, Kim JK, Marioni JC, Hore TA, Reik W. Genome-wide bisulfite sequencing in zygotes identifies demethylation targets and maps the contribution of TET3 oxidation. Cell Rep. 2014;9(6):1990–2000. doi:10.1016/j.celrep.2014.11.034.

Pfaffeneder T, Spada F, Wagner M, Brandmayr C, Laube SK, Eisen D, Truss M, Steinbacher J, Hackner B, Kotljarova O, Schuermann D, Michalakis S, Kosmatchev O, Schiesser S, Steigenberger B, Raddaoui N, Kashiwazaki G, Muller U, Spruijt CG, Vermeulen M, Leonhardt H, Schar P, Muller M, Carell T. Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. Nat Chem Biol. 2014;10(7):574–81. doi:10.1038/nchembio.1532.

Polanski Z, Motosugi N, Tsurumi C, Hiiragi T, Hoffmann S. Hypomethylation of paternal DNA in the late mouse zygote is not essential for development. Int J Dev Biol. 2008;52(2–3):295–8. doi:10.1387/ijdb.072347zp.

Popp C, Dean W, Feng S, Cokus SJ, Andrews S, Pellegrini M, Jacobsen SE, Reik W. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature. 2010;463(7284):1101–5. doi:10.1038/nature08829.

Probst AV, Santos F, Reik W, Almouzni G, Dean W. Structural differences in centromeric heterochromatin are spatially reconciled on fertilisation in the mouse zygote. Chromosoma. 2007;116(4):403–15. doi:10.1007/s00412-007-0106-8.

Ratnam S, Mertineit C, Ding F, Howell CY, Clarke HJ, Bestor TH, Chaillet JR, Trasler JM. Dynamics of Dnmt1 methyltransferase expression and intracellular localization during oogenesis and preimplantation development. Dev Biol. 2002;245(2):304–14. doi:10.1006/dbio.2002.0628.

Razin A, Webb C, Szyf M, Yisraeli J, Rosenthal A, Naveh-Many T, Sciaky-Gallili N, Cedar H. Variations in DNA methylation during mouse cell differentiation in vivo and in vitro. Proc Natl Acad Sci U S A. 1984;81(8):2275–9.

Reik W, Dean W, Walter J. Epigenetic reprogramming in mammalian development. Science. 2001;293(5532):1089–93. doi:10.1126/science.1063443.

Riggs AD, Jones PA. 5-methylcytosine, gene regulation, and cancer. Adv Cancer Res. 1983;40:1–30.

Rougier N, Bourc'his D, Gomes DM, Niveleau A, Plachot M, Paldi A, Viegas-Pequignot E. Chromosome methylation patterns during mammalian preimplantation development. Genes Dev. 1998;12(14):2108–13.

Russell GJ, Walker PM, Elton RA, Subak-Sharpe JH. Doublet frequency analysis of fractionated vertebrate nuclear DNA. J Mol Biol. 1976;108(1):1–23.

Saksouk N, Barth TK, Ziegler-Birling C, Olova N, Nowak A, Rey E, Mateos-Langerak J, Urbach S, Reik W, Torres-Padilla ME, Imhof A, Dejardin J, Simboeck E. Redundant mechanisms to form silent chromatin at pericentromeric regions rely on BEND3 and DNA methylation. Mol Cell. 2014;56(4):580–94. doi:10.1016/j.molcel.2014.10.001.

Sanford J, Forrester L, Chapman V, Chandley A, Hastie N. Methylation patterns of repetitive DNA sequences in germ cells of Mus musculus. Nucleic Acids Res. 1984;12(6):2823–36.

Santos F, Hendrich B, Reik W, Dean W. Dynamic reprogramming of DNA methylation in the early mouse embryo. Dev Biol. 2002;241(1):172–82. doi:10.1006/dbio.2001.0501.

Santos F, Peters AH, Otte AP, Reik W, Dean W. Dynamic chromatin modifications characterise the first cell cycle in mouse embryos. Dev Biol. 2005;280(1):225–36. doi:10.1016/j.ydbio.2005.01.025.

Santos F, Peat J, Burgess H, Rada C, Reik W, Dean W. Active demethylation in mouse zygotes involves cytosine deamination and base excision repair. Epigenetics Chromatin. 2013;6(1):39. doi:10.1186/1756-8935-6-39.

Seisenberger S, Andrews S, Krueger F, Arand J, Walter J, Santos F, Popp C, Thienpont B, Dean W, Reik W. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol Cell. 2012;48(6):849–62. doi:10.1016/j.molcel.2012.11.001.

Seki Y, Hayashi K, Itoh K, Mizugaki M, Saitou M, Matsui Y. Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. Dev Biol. 2005;278(2):440–58. doi:10.1016/j.ydbio.2004.11.025.

Seki Y, Yamaji M, Yabuta Y, Sano M, Shigeta M, Matsui Y, Saga Y, Tachibana M, Shinkai Y, Saitou M. Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. Development. 2007;134(14):2627–38. doi:10.1242/dev.005611.

Shirane K, Toh H, Kobayashi H, Miura F, Chiba H, Ito T, Kono T, Sasaki H. Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases. PLoS Genet. 2013;9(4):e1003439. doi:10.1371/journal.pgen.1003439.

Smallwood SA, Tomizawa S, Krueger F, Ruf N, Carli N, Segonds-Pichon A, Sato S, Hata K, Andrews SR, Kelsey G. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat Genet. 2011;43(8):811–4. doi:10.1038/ng.864.

Smith ZD, Chan MM, Mikkelsen TS, Gu H, Gnirke A, Regev A, Meissner A. A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature. 2012;484(7394):339–44. doi:10.1038/nature10960.

Szabo PE, Pfeifer GP. H3K9me2 attracts PGC7 in the zygote to prevent Tet3-mediated oxidation of 5-methylcytosine. J Mol Cell Biol. 2012;4(6):427–9. doi:10.1093/jmcb/mjs038.

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2009;324(5929):930–5. doi:10.1126/science.1170116.

Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006;126(4):663–76. doi:10.1016/j.cell.2006.07.024.

Tomizawa S, Kobayashi H, Watanabe T, Andrews S, Hata K, Kelsey G, Sasaki H. Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. Development. 2011;138(5):811–20. doi:10.1242/dev.061416.

Tsukada Y, Akiyama T, Nakayama KI. Maternal TET3 is dispensable for embryonic development but is required for neonatal growth. Sci Rep. 2015;5:15876. doi:10.1038/srep15876.

van der Heijden GW, Dieker JW, Derijck AA, Muller S, Berden JH, Braat DD, van der Vlag J, de Boer P. Asymmetry in histone H3 variants and lysine methylation between paternal and maternal chromatin of the early mouse zygote. Mech Dev. 2005;122(9):1008–22. doi:10.1016/j.mod.2005.04.009.

Vincent JJ, Huang Y, Chen PY, Feng S, Calvopina JH, Nee K, Lee SA, Le T, Yoon AJ, Faull K, Fan G, Rao A, Jacobsen SE, Pellegrini M, Clark AT. Stage-specific roles for tet1 and tet2 in DNA demethylation in primordial germ cells. Cell Stem Cell. 2013;12(4):470–8. doi:10.1016/j.stem.2013.01.016.

Wade PA, Gegonne A, Jones PL, Ballestar E, Aubry F, Wolffe AP. Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. Nat Genet. 1999;23(1):62–6. doi:10.1038/12664.

Wossidlo M, Arand J, Sebastiano V, Lepikhov K, Boiani M, Reinhardt R, Scholer H, Walter J. Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes. EMBO J. 2010;29(11):1877–88. doi:10.1038/emboj.2010.80.

Wossidlo M, Nakamura T, Lepikhov K, Marques CJ, Zakhartchenko V, Boiani M, Arand J, Nakano T, Reik W, Walter J. 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nat Commun. 2011;2:241. doi:10.1038/ncomms1240.

Wu H, Zhang Y. Reversing DNA methylation: mechanisms, genomics, and biological functions. Cell. 2014;156(1–2):45–68. doi:10.1016/j.cell.2013.12.019.

Xue JH, Xu GF, Gu TP, Chen GD, Han BB, Xu ZM, Bjoras M, Krokan HE, Xu GL, Du YR. Uracil-DNA Glycosylase UNG Promotes Tet-mediated DNA Demethylation. J Biol Chem. 2016;291(2):731–8. doi:10.1074/jbc.M115.693861.

Yamaji M, Ueda J, Hayashi K, Ohta H, Yabuta Y, Kurimoto K, Nakato R, Yamada Y, Shirahige K, Saitou M. PRDM14 ensures naive pluripotency through dual regulation of signaling and epigenetic pathways in mouse embryonic stem cells. Cell Stem Cell. 2013;12(3):368–82. doi:10.1016/j.stem.2012.12.012.

Young LE, Beaujean N. DNA methylation in the preimplantation embryo: the differing stories of the mouse and sheep. Anim Reprod Sci. 2004;82–83:61–78. doi:10.1016/j.anireprosci.2004.05.020.

Zheng P, Schramm RD, Latham KE. Developmental regulation and in vitro culture effects on expression of DNA repair and cell cycle checkpoint control genes in rhesus monkey oocytes and embryos. Biol Reprod. 2005;72(6):1359–69. doi:10.1095/biolreprod.104.039073.

Structure and Function of TET Enzymes

Xiaotong Yin and Yanhui Xu

Abstract

Mammalian DNA methylation mainly occurs at the carbon-C5 position of cytosine (5mC). TET enzymes were discovered to successively oxidize 5mC to 5-hydromethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC). TET enzymes and oxidized 5mC derivatives play important roles in various biological and pathological processes, including regulation of DNA demethylation, gene transcription, embryonic development, and oncogenesis. In this chapter, we will discuss the discovery of TET-mediated 5mC oxidation and the structure, function, and regulation of TET enzymes.

X. Yin
Fudan University Shanghai Cancer Center, Institute of Biomedical Sciences,
Shanghai Medical College of Fudan University, Shanghai 200032, China

Y. Xu (✉)
Fudan University Shanghai Cancer Center, Institute of Biomedical Sciences,
Shanghai Medical College of Fudan University, Shanghai 200032, China

Key Laboratory of Molecular Medicine, Ministry of Education,
Department of Systems Biology for Medicine, School of Basic Medical Sciences,
Shanghai Medical College of Fudan University, Shanghai 200032, China

State Key Laboratory of Genetic Engineering, Collaborative Innovation
Center of Genetics and Development, School of Life Sciences, Fudan University,
Shanghai 200433, China
e-mail: xuyh@fudan.edu.cn

© Springer International Publishing Switzerland 2016

A. Jeltsch, R.Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*,
Advances in Experimental Medicine and Biology 945,
DOI 10.1007/978-3-319-43624-1_12

Abbreviations

2HG	2-Hydroxyglutarate
5caC	5-Carboxylcytosine
5fC	5-Formylcytosine
5hmC	5-Hydromethylcytosine
5hmrC	5-Hydroxymethylcytidine
5mC	5-Methylcytosine
5mrC	5-Methylcytidine
6mA	N6-methyladenine
ABH2	AlkB homolog 2
AID	Activation-induced deaminase
AML	Acute myeloid leukemia
APOBEC	Apolipoprotein B mRNA-editing enzyme complex
BER	Base excision repair
CD	Catalytic domain
Chip-seq	Chromatin immunoprecipitation-sequencing
CMML	Chronic myelomonocytic leukemia
CpG	Cytosine-phosphate-guanine
CXXC	Cysteine-X-X-cysteine
Cys-C	Cys-rich C-terminal
Cys-N	Cys-rich N-terminal
Cys-rich	Cysteine rich
DMAD	DNA 6mA demethylase
DNMT	DNA methyltransferase
DSBH	Double-stranded β-helix
E11.5	Embryonic day 11.5
FH	Fumarate hydratase
HCF1	Host cell factor 1
HEK293	Human embryonic kidney 293
hmU	Hydroxymethyluracil
IDH	Isocitrate dehydrogenase
iPSCs	Induced pluripotent stem cells
JBP	J-binding protein
JmjC	Jumonji C
LC-MS	Liquid chromatography-mass spectrometry
MEFs	Mouse embryonic fibroblasts
mESCs	Mouse embryonic stem cells
MET	Mesenchymal to epithelial
NER	Nucleotide excision repair
NOG	N-oxalylglycine
OGT	O-linked β-N-acetylglucosamine transferase
OSKM	Oct4, Sox2, Klf4, and c-Myc
PGCs	Primordial germ cells

Pol II	RNA polymerase II
R-2HG	R-2-hydroxyglutarate
SAM	S-adenosyl methionine
SDH	Succinate dehydrogenase
SMUG1	Single-strand-selective monofunctional uracil DNA glycosylase 1
T7H	Thymine-7-hydroxylase
TAB-seq	Tet-assisted bisulfite sequencing
TCA	Tricarboxylic acid
TDG	Thymine-DNA glycosylase
TET	Ten-eleven translocation
TSKM	Tet1, Sox2, Kf4, and c-Myc
TSS	Transcription start site
α-KG	α-Ketoglutarate

1 Introduction

DNA methylation plays important roles in various biological processes through regulating gene expression, genome stability, genomic imprinting, and development (Bird 2002; Jaenisch and Bird 2003; Smith and Meissner 2013). Mammalian DNA methylation mainly occurs at the carbon-5 position of cytosine (5-methylcytosine, also known as 5mC) in CpG dinucleotide context (Bird 2002). Over the past several decades, mammalian DNA demethylation has been thought to be a passive process occurring through DNA replication-dependent dilution of the methylation mark in the absence of the maintenance DNA methyltransferase DNMT1. However, the passive dilution seems to be insufficient to account for the massive and cell division-independent DNA demethylation during specific stages of development. For example, genome-wide DNA demethylation in sperm-derived paternal pronuclei is detected shortly after fertilization before the completion of the first round of cell division (Mayer et al. 2000; Oswald et al. 2000). Proximal epiblast-derived primordial germ cells (PGCs) also undergo global DNA demethylation during their migration (Lee et al. 2002; Yamazaki et al. 2003; Hajkova et al. 2002). Considering that PGCs go through several cell cycles in the presence of DNMT1 before arriving at the genital ridge at E11.5, active DNA demethylation mechanisms are proposed to be involved in this process (Wu and Zhang 2010). Additionally, DNA demethylation has been reported to take place in the promoter region of different genes in somatic cells upon stimulation by certain signals, indicating the existence of active DNA demethylation (Bruniquel and Schwartz 2003; Kangaspeska et al. 2008; Martinowich et al. 2003; Metivier et al. 2008).

Over the past decades, a number of studies have reported the identification of different enzymes and pathways involved in DNA demethylation, including the enzymatic removal of the methyl groups, nucleotide excision repair (NER)

pathway, and deamination followed by base excision repair (BER) pathway. However, none of the above findings has been essentially proven. Recently, ten-eleven translocation (TET) proteins have been demonstrated to catalyze the oxidation of 5mC (Ito et al. 2011; He et al. 2011; Tahiliani et al. 2009). Further studies have indicated that TET and the oxidized 5mC derivatives play important roles in various biological and pathological processes, including the regulation of active DNA demethylation, gene transcription, embryonic development, and oncogenesis (Branco et al. 2012; Cimmino et al. 2011; Pastor et al. 2013; Tan and Shi 2012; Williams et al. 2012; Wu and Zhang 2011, 2014; Xu and Walsh 2014). Phylogenetic analyses show that members of the TET family are present in a vast variety of organisms from phages and fungus to plants and animals (Iyer et al. 2009, 2013). The prokaryotic TET proteins may generate 5hmC or hmU as an epigenetic mark or to help the prokaryotes to escape from the restriction-modification system of the host. Here, we will focus on the discovery of TET-mediated 5mC oxidation and the structure, function, and regulation of TET enzymes and briefly describe the subsequent steps in DNA demethylation.

2 Discovery of TET-Mediated 5mC Oxidation

2.1 TET-Mediated Iterative Oxidation of 5mC

Ten-eleven translocation 1 (TET1) was first identified as an MLL fusion partner in acute myeloid leukemia (AML) patients (Ono et al. 2002). Mammalian TET proteins have three family members, TET1, TET2, and TET3. TET proteins were considered as candidates in a search for 5mC-modifying enzymes in mammals (Tahiliani et al. 2009), because they show homology to the trypanosome proteins J-binding protein 1 (JBP1) and JBP2, which have been proposed to possess hydroxylase activity toward the C5-methyl group of thymine (Cliffe et al. 2009; Yu et al. 2007). TET1 has been identified to catalyze the hydroxylation of 5mC to generate 5-hydroxylmethylcytosine (5hmC) in a manner dependent on α-ketoglutarate (α-KG) and Fe(II) (Tahiliani et al. 2009). This activity has been observed for all three mouse Tet enzymes (Ito et al. 2010).

Later on, two groups independently demonstrated that TET mediate iterative oxidation of 5mC to 5hmC, 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) (He et al. 2011; Ito et al. 2011), reminiscing the thymine-7-hydroxylase (T7H)-catalyzed stepwise conversion of thymine to isoorotate (Smiley et al. 2005; Neidigh et al. 2009; Liu et al. 1973). The three 5mC oxidation derivatives counteract the DNMT1-mediated maintenance methylation, because CpG sites containing oxidized 5mC are not good substrates for DNMT1 (Hashimoto et al. 2012; Valinluck and Sowers 2007). Following studies have demonstrated that active DNA demethylation is dependent on TET3 in the early stage of embryo development (Guo et al. 2014; Shen et al. 2014a; Gu et al. 2011). Thus, TET enzymes may facilitate DNA demethylation through passive dilution of the modified bases during replication and/or via active demethylation (Fig. 1).

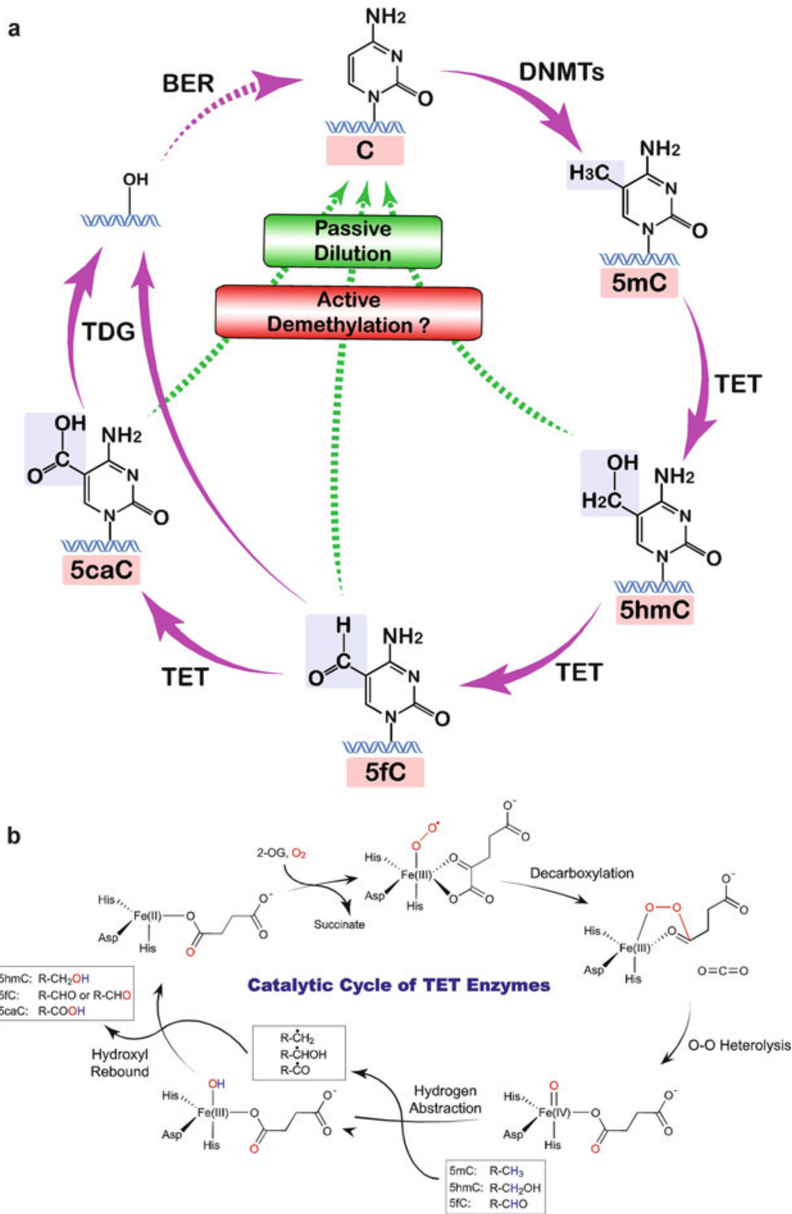


Fig. 1 Pathways for dynamic DNA methylation. (**a**) Mammalian DNA methylation (5mC) is established by de novo DNA methyltransferases DNMT3A/3B, and the patterns of 5mC are maintained by DNMT1. TET enzymes successively oxidize 5mC to 5hmC, 5fC, and 5caC. The three 5mC derivatives may be involved in DNA demethylation through passive dilution and active demethylation. The latter has been observed in early embryo development, but the mechanism remains unknown. 5fC/5caC is replaced by unmodified cytosine through the TDG-initiated BER pathway, which does not account for massive DNA demethylation in mouse zygotes but contributes to demethylation of specific loci in ESCs and neurons. (**b**) Schematic model for the oxidative reactions catalyzed by TET enzymes

2.2 TET-Dependent DNA Demethylation

An additional step for the active DNA demethylation should exist because 5mC and its oxidation derivatives 5hmC/5fC/5caC do not spontaneously convert to unmodified cytosine under physiological conditions. Interestingly, 5fC and 5caC can be recognized and excised by thymine-DNA glycosylase (TDG) coupled with the BER pathway (He et al. 2011; Maiti and Drohat 2011; Zhang et al. 2012) to complete the demethylation pathway (Fig. 1). However, TDG is not required for active demethylation during early stage embryo development (Guo et al. 2014). In the thymidine salvage pathway, the thymine to uracil conversion is achieved by thymine hydroxylation to isoorotate, followed by decarboxylation catalyzed by the isoorotate decarboxylase (Neidigh et al. 2009; Smiley et al. 2005; Xu et al. 2013). Decarboxylation has been proposed to contribute to the conversion of 5caC to an unmodified cytosine. In support of this hypothesis, a 5caC-decarboxylation activity has been observed in nuclear extracts of mouse embryonic stem cells (mESCs) (Schiesser et al. 2012). However, to what extent this decarboxylation activity contributes to the active demethylation remains largely unknown.

5mC derivatives have been proposed to serve as intermediates in DNA demethylation through alternative pathways. For example, cytidine deaminase AID (activation-induced deaminase)/APOBEC (apolipoprotein B mRNA-editing enzyme complex) family members have been proposed to deaminate 5hmC to generate 5hmU, which is further removed by SMUG1 (single-strand-selective monofunctional uracil DNA glycosylase 1) or TDG and ultimately replaced by cytosine via the BER pathway (Guo et al. 2011). However, controversial evidence for this hypothesis does exist. For example, 5hmU is not detectable in HEK293T cells concomitantly with the overexpression of AID/APOBEC. Purified AID/APOBEC possess decreased or undetectable deamination activity toward 5mC or 5hmC compared with the canonical substrate cytosine, because of steric effects of the 5-substituent groups (Nabel et al. 2012; Rangam et al. 2012). Intriguingly, DNMT3A/3B have been shown to directly convert 5hmC to cytosine in the absence of S-adenosyl-L-methionine (AdoMet) under certain reaction conditions (Liutkeviciute et al. 2009; Chen et al. 2012). Moreover, a decarboxylation activity toward 5caC mediated by bacterial and mammalian DNA methyltransferases, such as M.HhaI and DNMTs, was reported (Liutkeviciute et al. 2014). However, whether this reaction occurs in vivo remains elusive (Shen et al. 2014b), because AdoMet acts as a general methyl donor in various biological processes and it is very abundant in cells.

2.3 Mechanisms for TET-Mediated 5mC Oxidation

TET enzymes belong to the group of α-KG/Fe(II)-dependent dioxygenases (Tahiliani et al. 2009; Iyer et al. 2009) (Fig. 2). Members of this enzyme family regulate secondary metabolisms in plants and microorganisms, biosynthesis of collagen, hypoxia response, and epigenetic modification in animals, through catalyzing versatile oxidative reactions, such as hydroxylation, desaturation, epoxidation, epimerization, and oxidative halogenations (McDonough et al. 2010; Loenarz and

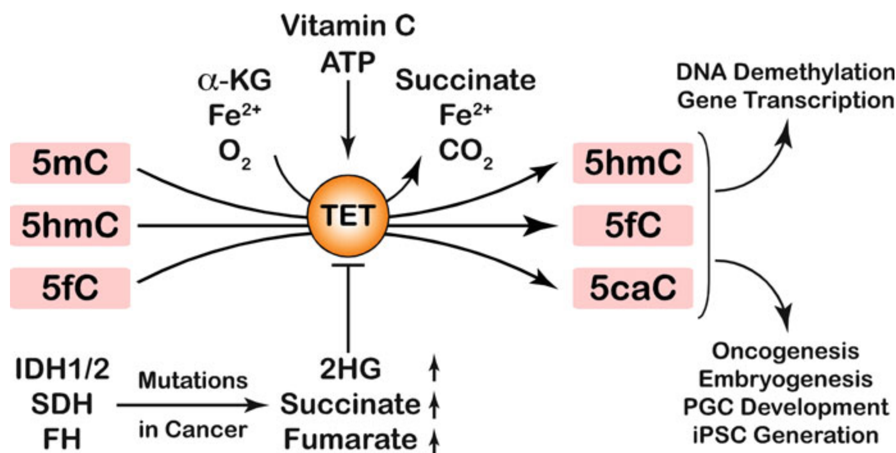


Fig. 2 TET-mediated oxidation of 5mC. TET enzymes successively oxidizes 5mC to 5hmC, 5fC, and 5caC. The oxidation depends on the presence of Fe(II), molecular oxygen, and α-KG as co-substrates. Vitamin C and ATP enhance the enzymatic activity of TET. TET enzymes and 5mC oxidation derivatives are involved in various molecular events (DNA demethylation and regulation of gene expression) and therefore play key roles in embryonic development and oncogenesis. Patient-derived mutations of enzymes in TCA cycle, including IDH1/2, SDH, and FH lead to accumulation of metabolites 2-HG, succinate, and fumarate. These α-KG analogs competitively inhibit the activities of TET enzymes and JmjC-containing histone demethylases and therefore may contribute to oncogenesis

Schofield 2008). In mammals, the group includes the AlkB family of DNA/RNA demethylases, JmjC-containing histone lysine demethylases, TET enzymes, and other enzymes that will not be discussed hereafter.

Oxidation mediated by the α-KG/Fe(II)-dependent dioxygenases can be divided into two successive steps, dioxygen activation and substrate oxidation (Fig. 1b) (Shen et al. 2014b). In the first step, Fe(II) and α-KG coordinate to the "facial triad" composed of the conserved HxD/E…H (where x can be any residue) motif of the enzyme. Then substrate deposition allows molecular dioxygen to replace the water molecule and bind to the Fe(II) in the catalytic center of the enzyme. One oxygen atom of the bound dioxygen inserts into the succinate derived from α-KG decarboxylation, and the other one couples with the iron to generate a high-valent Fe(IV)-oxo intermediate (Seisenberger et al. 2012; Krebs et al. 2007; Valegard et al. 2004). The dioxygen activation procedure has been demonstrated to be accelerated in the presence of substrate within the catalytic cavity (Ryle et al. 1999), which induces environmental changes in Fe(II) coordination and then facilitates dioxygen binding (Muller et al. 2005). In the second step, the C-H bond of the substrate is cleaved by the Fe(IV)-oxo oxidizing radical, and the oxygen atom is transferred to the target carbon group through hydrogen abstraction (Price et al. 2003; Hoffart et al. 2006). Upon substrate oxidation, the iron returns to the Fe(II) state and thus completes one reaction cycle. TET enzymes oxidize 5mC through the same mechanism (Fig. 2) (Shen et al. 2014b). In the first cycle of oxidation, 5mC is converted to 5hmC. In the next step, 5hmC is further oxidized to a germinal diol, which decomposes into 5fC,

and then 5caC is generated in a third cycle. One α-KG is consumed for each cycle of the reaction (Fig. 1b).

2.4 Oxidation of 5mrC-RNA and 6mA-DNA

In addition to their activity toward 5mC in DNA, enzymes of the TET family also possess oxidation activity on other substrates, including 5-methylcytidine (5mrC) in RNA and N6-methyladenine (6mA) in DNA. Fu et al. found that TET enzymes convert 5mrC to 5-hydroxymethylcytidine (5hmrC) in RNA in vitro and induce the formation of 5hmrC in human cells (Fu et al. 2014). The presence of 5hmrC in RNA has been verified in mammalian cells using sensitive and accurate LC-MS/MS/MS. This study has given rise to a number of open questions. For example, can 5hmrC-RNA be further oxidized similarly to 5hmC in DNA? What is the function of TET in mediating the oxidation of 5mrC in RNA? Additionally, 5mrC in RNA should be recognized by TET enzymes in a manner similar to that of 5mC in DNA during oxidation. Consequently, how do TET enzymes recognize 5mrC-RNA?

DNA 6mA is commonly found in bacterial genomes. Recent studies have demonstrated the existence of this modified nucleotide in genomes of various eukaryotes, including *Chlamydomonas* (Fu et al. 2015), *C. elegans* (Greer et al. 2015), and *Drosophila* (Zhang et al. 2015). While searching for a specific enzyme responsible for demethylation of 6mA, Zhang et al. found a *Drosophila* gene (CG2083) that possesses such activity and named the corresponding protein DMAD (DNA 6mA demethylase), which is a homolog of mammalian TET (Zhang et al. 2015). Nuclear extracts from late-stage embryos showed considerable 6mA demethylation activity, whereas depletion of DMAD from the nuclear extracts using an anti-DMAD antibody or siRNA led to the loss of 6mA demethylase activity. Further study has suggested that DMAD removes 6mA primarily from transposon regions and is essential for development (Zhang et al. 2015). In an in vitro assay, the catalytic domain (CD) of DMAD shows 5mC oxidation activity, albeit approximately 30-fold lower than that of Tet1-CD, suggesting dual substrate specificity. This study is the first report to indicate that TET enzymes can oxidize nucleotides other than 5mC/5mrC. Intriguingly, structure-based sequence analysis has indicated that all critical residues for 5mC recognition are conserved in human TET2 and *Drosophila* DMAD (Hu et al. 2013). Because 6mA is considerably larger than 5mC, it would be of interest to investigate how 6mA is bound and specifically recognized by DMAD.

3 Function of TET Enzymes

3.1 Distribution of TET Enzymes and 5mC Oxidation Derivatives

The level of 5hmC is relatively high in neurons (15–40 % of 5mC), and self-renewing and pluripotent stem cells, but is greatly reduced in cancer cells and along with differentiation (Globisch et al. 2010; Ito et al. 2011; Mellen et al. 2012;

Pfaffeneder et al. 2011; Song et al. 2011; Tahiliani et al. 2009; Szwagierczak et al. 2010; Ruzov et al. 2011), suggesting a positive correlation between 5hmC and pluripotent states. Based on this finding 5hmC might serve as not only a DNA demethylation intermediate but also a relatively stable epigenetic mark. The 5mC derivatives may reduce DNA-binding affinity of methyl-CpG-binding proteins (Valinluck et al. 2004) or could be recognized by other chromatin-associated factors for transcriptional regulation (Frauer et al. 2011; Yildirim et al. 2011; Zhou et al. 2014).

Consistent with the enrichment of 5hmC in mESCs, relatively high expression levels of Tet1 and Tet2 (to a lesser extent Tet3) are also detected in mESCs and mouse iPSCs, whereas the protein levels diminish during mESCs differentiation (Ito et al. 2010; Szwagierczak et al. 2010; Wossidlo et al. 2011; Ficz et al. 2011). TET enzymes are expressed at different magnitudes in adult human and mouse tissues (Ito et al. 2010). In particular, Tet1 is largely expressed in PGCs (Yamaguchi et al. 2012). Tet3 is highly expressed in oocytes and zygotes, whereas Tet1 or Tet2 shows undetectable or moderate expression (Wossidlo et al. 2011), indicating a critical role for Tet3 in epigenetic reprogramming during embryo development. The expression levels of TET enzymes are largely decreased in various cancers which is also consistent with the low level of 5hmC (Lian et al. 2012; Yang et al. 2013; Haffner et al. 2011).

In accordance with other chromatin-modifying enzymes, overexpressed TET1 is observed in the nucleus, but not the cytoplasm in cultured HEK293 cells by immunofluorescence staining (Tahiliani et al. 2009). Transfected Tet1, Tet2, and Tet3 are also found localized in the nucleus in cultured U2OS and HEK293T cells by another group (Ito et al. 2010). However, Tet3 is found present in the male pronucleus in the zygotic stage, and it translocates to the cytoplasm in the preimplantation stages (Gu et al. 2011). Besides, Aid may regulate the subcellular localization of Tet proteins. When co-transfected with Aid, the subcellular localization of Tet1/2/3 is altered to the cytoplasm from the nucleus, and this translocation is associated with Aid shuttling (Arioka et al. 2012).

A base resolution map of 5hmC in human and mouse ESCs determined by Tet-assisted bisulfite sequencing (TAB-seq) has revealed its widespread distribution in the genome and varied abundance across functional elements (Yu et al. 2012). Almost half (46.4%) of the 5hmCs reside in distal regulatory elements, in which p300-binding sites, enhancers, CTCF-binding sites, and DNaseI hypersensitive sites harbor more 5hmC than other genic regions. High levels of 5hmC concomitant with low levels of 5mC have been observed near the transcription factor binding sites but not within them. 5mC but not 5hmC is enriched in repetitive elements. Notably, TAB-seq analysis has indicated that 5hmC is abundant in regions with low CpG content (Yu et al. 2012), whereas antibody-based sequencing has indicated that 5hmC is enriched in CpG-rich transcription start sites (TSSs) (Williams et al. 2011). This discrepancy may result from the different approaches used.

Genome-wide analysis of 5fC localization in mESCs has revealed that 5fC is enriched at poised and active enhancers, preferentially at poised enhancers, suggesting its roles in epigenetic priming (Song et al. 2013). An additional study using DNA immunoprecipitation-coupled chemical modification-assisted

bisulfite sequencing has indicated that 5fC and 5caC can be detected in enhancers, promoters, and intragenic regions but share limited overlap in mESCs genome (Lu et al. 2015). In addition, the 5mC oxidation activity negatively correlates with 5mC abundance and positively correlates with enhancer activity, suggesting its roles in regulating gene expression (Lu et al. 2015) by affecting the process of transcriptional elongation. In support of this hypothesis, *Tdg*-KO mESCs display an apparent retardation of RNA polymerase II (Pol II) elongation compared with the wild-type cells, possibly because of the increased level of 5caC. The crystal structure of Pol II in complex with 5caC-DNA indicates that an extra hydrogen bond formed between Pol II and 5caC may contribute to this retardation (Wang et al. 2015a). These studies suggest a functional impact of 5caC on gene expression.

3.2 Tet1 in mESCs and Cell Differentiation

Individual knockdown of *Tet1* or *Tet2* leads to a partial reduction of 5hmC level, suggesting that both TET1 and TET2 are responsible for 5hmC maintenance in human and mouse ESCs (Dawlaty et al. 2013; Koh et al. 2011; Tahiliani et al. 2009). Chip-seq analyses have revealed that most of the Tet1 binding sites co-localize with 5hmC in euchromatin regions in mESCs and strongly accumulate at hypomethylated CpG-rich promoters (Williams et al. 2011; Wu et al. 2011). *Tet1* knockdown or knockout in mESCs gives rise to increased DNA methylation levels, suggesting that Tet1 is required to maintain the hypomethylation state of many gene promoters (Wu et al. 2011). Tet1 regulates the expression of genes related to the maintenance of mESCs, through promoting the transcription of pluripotency-associated factors and suppressing the expression of differentiation-associated factors (Dawlaty et al. 2011; Ficz et al. 2011; Williams et al. 2011; Wu et al. 2011).

Although Tet1 plays a role in gene regulation, it is largely dispensable in the maintenance of pluripotency; however, it may regulate the cell lineage differentiation. *Tet1* knockdown and knockout mESCs present an unvaried propagation rate and morphology (Ficz et al. 2011; Dawlaty et al. 2011; Koh et al. 2011). The full pluripotency of *Tet1*$^{-/-}$ mESCs has also been verified by the tetraploid complementation assay (Dawlaty et al. 2011). *Tet1* knockdown or knockout mESCs generate teratomas that contain differentiated cells from all three germ layers but show skewed differentiation toward the trophectoderm (Dawlaty et al. 2011; Koh et al. 2011). Consistently, *Tet1* knockdown mESCs show increased mRNA levels of trophectoderm markers *Cdx2*, *Eomes*, and *Hand1* but decreased mRNA levels of *Pax6* and *Neurod1*, which are representatives of neuroectoderm (Koh et al. 2011). The *Tet1* and *Tet2* double-knockout mESCs retain pluripotency in the teratoma formation assay but generate abnormal chimeric embryos (Dawlaty et al. 2013). Although there exist possible defects in embryo development as indicated by the significant skewing, *Tet1*$^{-/-}$ mice are viable and fertile and grow normally, except for the smaller size of some of the pups (Dawlaty et al. 2011), and viable *Tet1* and *Tet2* double-knockout mice can also be generated, albeit with a low birthrate (Dawlaty et al. 2013).

3.3 TET3 in Early Embryogenesis

Active demethylation of the paternal pronucleus occurs rapidly after fertilization (Mayer et al. 2000; Oswald et al. 2000). Recent studies have suggested that TET-mediated 5mC oxidation participates in this demethylation process (Wossidlo et al. 2011; Iqbal et al. 2011; Gu et al. 2011). Immunofluorescence staining has indicated that 5mC of the maternal pronucleus remains constant through different pronucleus stages, whereas 5mC in paternal pronuclei is progressively lost and 5hmC shows up. Coincident with the appearance of 5hmC, Tet3 is especially enriched in oocytes and paternal pronuclei of zygotes (Wossidlo et al. 2011; Gu et al. 2011; Iqbal et al. 2011), suggesting its key role in DNA demethylation. Tet3-deficient zygotes show impaired conversion of 5mC to 5hmC in paternal genome and retardation of demethylation of paternal *Oct4* and *Nanog* genes, indicating that Tet3 plays an important role in epigenetic reprogramming (Gu et al. 2011). Tet3 also generates 5fC and 5caC in paternal pronuclei (Inoue et al. 2011), which may also contribute to DNA demethylation (Fig. 1).

It is generally believed that 5mC in maternal pronuclei does not undergo active demethylation (Inoue et al. 2011; Inoue and Zhang 2011; Iqbal et al. 2011; Gu et al. 2011; Xie et al. 2012). The corresponding 5mC seems to be protected by PGC7 (also known as Dppa3 or Stella) (Nakamura et al. 2007; Wossidlo et al. 2011; Iqbal et al. 2011). However, advanced single-base resolution sequencing analyses of mouse oocytes and early embryos have indicated that both maternal and paternal pronuclei undergo active demethylation and replication-dependent dilution of methylation (Wang et al. 2014; Guo et al. 2014; Shen et al. 2014a). Wang et al. have demonstrated the existence of 5hmC and 5fC in both male and female pronuclei and showed that a significant proportion of maternal genomes are subjected to active demethylation (Wang et al. 2014). Studies from another two groups have indicated that replication-dependent dilution contributes more than expected to genome-wide demethylation in both maternal and paternal pronuclei (Guo et al. 2014; Shen et al. 2014a). Nevertheless, active DNA demethylation does occur in a manner dependent on TET3, although the mechanism remains elusive (Fig. 1).

3.4 TET Enzymes in PGC Development

During the development of PGCs, massive genome-wide demethylation occurs for the establishment of totipotency of the germ cells (Saitou et al. 2012). *Tet1* (but not *Tet2* or *Tet3*) is substantially expressed in mouse PGCs and is enriched between E10.5 and E11.5, suggesting its key role in regulating methylation states of imprinting genes during PGCs development (Hackett et al. 2013; Yamaguchi et al. 2012, 2013). *Tet1* paternal knockout mice have been obtained by mating *Tet1*$^{-/-}$ male and wild-type female mice, but the mice exhibit partial fetal or postnatal defects as well as early embryo lethality (Yamaguchi et al. 2013).

Interestingly, 5mC and 5hmC diminish at different rates during PGCs development (Hackett et al. 2013), suggesting the occurrence of two waves of genomic demethylation. During the first wave, 5mC drops to approximately 30 % until E9.5 during PGC migration (Seisenberger et al. 2012) owing to passive dilution

(Kurimoto et al. 2008). During the second wave, active demethylation occurs in a manner dependent on TET1 and possibly TET2, leading to a rapid decrease in 5mC and a slight increase in 5hmC, followed by a decline resembling a dilution pattern (Hackett et al. 2013; Yamaguchi et al. 2013).

3.5 TET Enzymes in Somatic Cell Reprogramming

In induced pluripotent stem cells (iPSCs) from mouse embryonic fibroblasts (MEFs) induced by OSKM factors (Oct4, Sox2, Klf4, and c-Myc), Tet1 has a high expression level, whereas Tet2 expression is relatively moderate and Tet3 is undetectable (Gao et al. 2013). MEFs lacking all three Tet enzymes no longer produce reprogrammed colonies; this result is in contrast to MEFs deficient in one or two of the Tet enzymes, indicating that Tet enzymes are redundant but essential for iPSCs generation (Hu et al. 2014). Furthermore, Tet1 overexpression can facilitate OSKM (Oct4, Sox2, Klf4, and c-Myc)-induced iPSC formation (Gao et al. 2013). The underlying mechanisms for Tet functions in iPSC generation have been studied from various aspects, as indicated below.

Nanog and *Esrrb*-included pluripotency loci require reactivation during somatic cell reprogramming. Tet2 is recruited to these pluripotency loci and may contribute to transcriptional induction in the early stage of reprogramming (Doege et al. 2012). Another study has shown the physical association and synergetic effect of Nanog and TET1 or TET2 in the enhancement of iPSCs generation. Nanog deficiency leads to reduced recruitment of TET1 toward a subset of genomic loci shared by Nanog and TET1, suggesting that Nanog recruits TET1 to target genes for pluripotency establishment and cell lineage specification (Costa et al. 2013). Intriguingly, somatic cell reprogramming can be induced by TSKM factors (Tet1, Sox2, Klf4, and c-Myc) (Gao et al. 2013). In this system, Oct4 in OSKM factors (Oct4, Sox2, Klf4, and c-Myc) is replaced by Tet1, which facilitates iPSC generation by promoting the demethylation and reactivation of Oct4.

Another study has shown that the depletion of all three Tet enzymes prevents iPSC generation, because the mesenchymal-to-epithelial transition (MET) step is blocked. This barrier could be ascribed to the inactivation of a cluster of miRNAs (Hu et al. 2014). These miRNAs belong to the miR-200 family and are known to modulate the expression of transcription factors that inhibit the epithelial markers. Tet and TDG are required for the demethylation and reactivation of miR-200 and for the restoration of MET and the initiation of reprogramming process. Reintroducing only this cluster of miRNAs overcomes the barrier and completes the reprogramming, suggesting the nonessential function of Tet enzymes in the subsequent processes.

3.6 TET Enzymes and Cancer

TET2 mutations are frequently observed in chronic myelomonocytic leukemia (CMML), AML, and other myeloid malignancy patients (Abdel-Wahab et al. 2009; Langemeijer et al. 2009; Tefferi et al. 2009). Some of the patient-derived mutations,

such as those involved in iron chelation (H1881, H1382, and D1384), α-KG interaction (R1896, R1261, and S1898), and DNA recognition (N1387, H1904, and Y1902), largely decrease or abolish TET2 activity (Ko et al. 2010; Hu et al. 2013). The prevalence of two catalytic inactive mutants (H1802 and R1817, corresponding to H1881 and R1896 in human TET2) strongly correlates with low genomic 5hmC levels in the bone marrow and blood. Further evidence has shown that TET2 is critical for normal myelopoiesis (Ko et al. 2010). *Tet2* deficiency results in enhanced self-renewal and the abnormal proliferation of hematopoietic stem cells, resulting in splenomegaly, monocytosis, and extramedullary hematopoiesis in an animal model (Moran-Crusio et al. 2011). *Tet2*$^{+/-}$ mice are also predisposed for myeloid transformation (Moran-Crusio et al. 2011). *TET2* mutations have been identified in human lymphomas, and these loss-of-function mutants may perturb the early developmental state of hematopoietic stem cells, leading to myeloid and/or lymphoid malignancies (Quivoron et al. 2011).

Melanoma is a highly malignant and aggressive type of cancer and displays globally hypomethylation and gene-specific hypermethylation at certain tumor suppressors (Hoon et al. 2004; Liu et al. 2008; Shen et al. 2007). 5hmC is considerably decreased in melanoma, compared with melanocytes and nevi, suggesting that the modification may be a unique feature of melanoma. Moreover, all three TET proteins are downregulated in melanoma, which is consistent with the decreased 5hmC level (Lian et al. 2012). The low level of 5hmC and downregulation of TET enzymes have also been found in other human cancer cells, including breast, liver, lung, pancreatic, colon, and prostate cancers (Yang et al. 2013; Haffner et al. 2011).

3.7 TET Enzymes in Neural System

The presence of 5hmC in mammalian genomes was first discovered in Purkinje neurons and the brain (Kriaucionis and Heintz 2009). Relatively high levels of 5hmC have been observed in various adult brain regions (Munzel et al. 2010; Ruzov et al. 2011; Szwagierczak et al. 2010), suggesting important roles for TET enzymes and 5mC oxidation derivatives in neural systems. TET1 exhibits strong co-localization with the neuronal marker NeuN throughout the hippocampus, implicating a primary expression and distribution in neurons (Kaas et al. 2013). Overexpression of TET1 leads to increased conversion of 5mC to 5hmC in the central nervous system and TET1 is essential for the demethylation of fibroblast growth factor 1 (*Fgf1*) and brain-derived neurotrophic factor (*Bdnf*) promoters (Guo et al. 2011).

Tet1 knockout mice display normal overall health and brain development (Rudenko et al. 2013; Gao et al. 2013). However, depletion of Tet1 leads to a hypermethylation and downregulation of genes involved in progenitor cells proliferation and therefore impaired hippocampal neurogenesis and the mice exhibit poor learning and memory (Gao et al. 2013). Another group reported that the Tet1 ablation gives rise to the downregulation of genes involved in neuronal activity and results in impairment of memory extinction, synaptic plasticity, and hippocampal long-term depression (Rudenko et al. 2013).

4 Structure of TET Enzymes

4.1 Domain Structure of Human TET Enzymes

The three TET family members (TET1, TET2, and TET3) share a less conserved N-terminal region and highly conserved C-terminal catalytic domain, a characteristic fold in α-KG and Fe(II)-dependent dioxygenases (Fig. 3a). TET1 (also known as CXXC6) and TET3 (also known as CXXC10) contain an N-terminal zinc finger cysteine-X-X-cysteine (CXXC) domain, which was previously thought to recognize unmethylated CpGs (Long et al. 2013). Intriguingly, *TET2* does not encode a CXXC domain, but in the genome it is located close to the *IDAX* gene, which encodes a CXXC domain similar to that of TET1 and TET3. IDAX directly interacts with TET2 and is enriched at unmethylated CpGs (Ko et al. 2013). Genome-wide analysis in mESCs indicates that Tet1 preferentially localizes to the TSS of unmethylated CpG-rich promoters and within genes (Williams et al. 2011; Wu et al. 2011). Biochemical analysis indicates that the CXXC domain of Tet1 binds CpG and methylated-CpG DNA (Xu et al. 2011b). A further structural study has indicated that the Tet3 CXXC domain prefers unmethylated cytosines within CpG or non-CpG DNA, and the CXXC domain is critical for Tet3 targeting (Xu et al. 2012). Thus, the CXXC domain of TET enzymes may recognize CpG-containing DNA and accommodate cytosine methylation, thereby providing flexibility for their genomic targeting. Therefore, it is of interest to investigate how the TET CXXC domains recognize CpG DNA when the cytosine is replaced by 5hmC, 5fC, and 5caC.

A previous study has predicted that TET enzymes contain a double-stranded β-helix (DSBH) fold, which is a characteristic domain of α-KG/Fe(II)-dependent dioxygenases and a cysteine-rich (Cys-rich) domain at the N-terminus of the DSBH (Iyer et al. 2009). Both the DSBH and Cys-rich domains are highly conserved among TET enzymes and across species. There is a much less conserved low-complexity insert within the core DSBH domain. The deletion of the insert does not obviously affect the in vitro enzymatic activity of TET2 (Hu et al. 2013). However, the insert is present across the entire TET family enzymes, suggesting that it might be important for TET functions in vivo.

4.2 Crystal Structure of the TET2-5mC-DNA Complex

The function of TET enzymes has been extensively studied since the discovery of the TET-mediated 5mC oxidation in 2009 (Branco et al. 2012; Cimmino et al. 2011; Pastor et al. 2013; Tan and Shi 2012; Williams et al. 2012; Wu and Zhang 2011, 2014; Xu and Walsh 2014). However, a few fundamental questions remain to be addressed. For example, how do TET enzymes specifically recognize their DNA substrate? How do TET enzymes successively oxidize 5mC to 5hmC, 5fC, and 5caC? Finally, how do patient-derived mutations of TET2 affect its enzymatic activity and contribute to oncogenesis? The three-dimensional structure of TET enzymes in complex with DNA substrate will provide valuable information to address these questions.

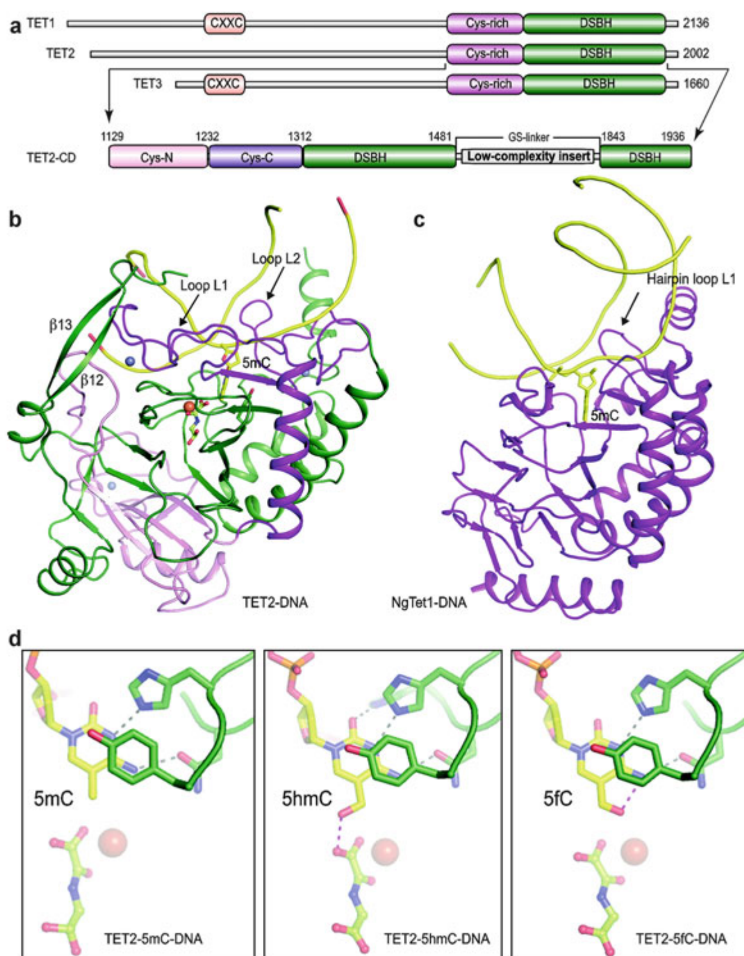


Fig. 3 Structures of TET enzymes in complex with DNA substrates (**a**) Color-coded domain architecture of human TET enzymes. All three TET family members share conserved Cys-rich and DSBH domains, which constitute the catalytic domain and are both essential for enzymatic activities of TET enzymes. TET1 and TET3 contain a CXXC domain, which recognizes CpGs and is essential for the function of TET enzymes in vivo. The TET2 construct used for structural study is indicated below. (**b**) Crystal structure of human TET2 in complex with methylated DNA. The color scheme is used as in Fig. 3a. The DNA is shown in ribbon representation and colored in *yellow*. NOG, the α-KG analog, is shown in stick representation. Fe(II) and zinc cations are shown as *red* and *gray* balls, respectively. (**c**) Ribbon representation of the crystal structure of NgTet1-DNA complex. The structure is shown in a similar orientation as that of TET2 in Fig. 3b for comparison. (**d**) Structural comparison of TET2-5mC-DNA (PDB: 4NM6, 2.02Å resolution), TET2-5hmC-DNA (PDB: 5DEU, 1.80 Å resolution), and TET2-5fC-DNA (PDB: 5D9Y, 1.97Å resolution) complexes (Hu et al. 2013, 2015). The three complexes adopt similar overall structures (not shown). The close-up views of the TET2-DNA interactions show the different conformation of 5mC/5hmC/5fC within the catalytic cavity of TET2. Critical bases or residues are shown in stick representation. Hydrogen bonds are indicated as *dashed lines*. The nitrogen, oxygen, and phosphorous atoms are shown in *blue*, *red*, and *orange*, respectively. All the structural figures were modified from published literatures (Hu et al. 2013, 2015; Hashimoto et al. 2014)

The challenge for structural studies of TET enzymes is the difficulty in obtaining well-behaved protein for crystallization. Hu et al. have mapped the minimal regions required for the enzymatic activity of TET2 (Hu et al. 2013). As shown in Fig. 3a, a human TET2 construct corresponding to residues 1129–1936 with an internal deletion (residues 1481–1843) maintains enzymatic activity and largely improves the solubility and yield of protein expression and purification. The TET2-5mC-DNA complex structure has been determined with this fragment at 2.02 Å resolution (Hu et al. 2013) (Fig. 3b). The α-KG analog N-oxalylglycine (NOG) was used to avoid the oxidation of 5mC in the crystals. Notably, the TET2 regions in the crystal structure are highly conserved in most TET enzymes, indicating that the structural features described below apply to other TET enzymes.

The structure shows a compact fold of the catalytic domain in complex with the 5mC-DNA duplex. The central DSBH core is formed by two β-sheets and stabilized by other regions from both sides and on the bottom. The Cys-rich domain wraps around the DSBH core and is separated into Cys-rich N-terminal (Cys-N) and C-terminal (Cys-C) subdomains. Two β-strands (β12 and β13) should be connected by the large low-complexity insert, which has been removed for crystallization (Fig. 3b). This organization is consistent with the observation that the insert is not required for TET activity and suggests that the insert is located away from the central catalytic domain of TET2.

The overall structure of TET2 is further stabilized by three zinc cations. Notably, Zn2 and Zn3 are coordinated by residues from both Cys-rich and DSBH domains and thus bring flexible regions from the two domains together to facilitate the overall structure formation. Sequence analysis has indicated that all of the residues involved in zinc coordination are highly conserved, suggesting that this architecture is critical for TET enzymes. The Fe(II) and NOG are localized in the center of the DSBH core domain and are bound and stabilized by highly conserved residues. Mutation of these residues significantly decreases or abolishes the enzymatic activity of TET2 (Hu et al. 2013; Ko et al. 2010). Most of the patient-derived mutations occur at the residues for zinc and iron coordination or the α-KG interaction. However, there are quite a number of mutations occurring at residues that do not directly impair TET2 activity. For example, some of the mutations are at residues within the insert region or the N-terminus. Whether and how these mutations contribute to oncogenesis requires further investigation.

The methylated dsDNA is located above the DSBH core and is stabilized by two loops (L1 and L2) from Cys-C subdomain (Fig. 3b). One methylcytosine (5mC) is flipped out of the DNA duplex and inserted into the catalytic cavity. As a replacement, a hydrophobic loop fills in this gap within the double-stranded DNA. A highly conserved residue (Y1294) stabilizes the G:C base pair of the CpG dinucleotide through a base-stacking interaction. Therefore, TET2 may use a tipping mechanism to search for the modified CpG when DNA slides through the enzyme (Tsai and Tainer 2013). The mechanism is similar to that observed for AlkB homolog 2 (ABH2) (Yang et al. 2008, 2009). The comparison between TET2-DNA and ABH2-DNA structures has been described previously (Hu et al. 2013; Tsai and Tainer 2013) and will not be discussed here.

Within the catalytic cavity, the 5mC is specifically recognized by TET2 through a network of interactions, which allow the 5mC to adopt a specific orientation so that the methyl group faces toward the catalytic center for reaction (Hu et al. 2013). The catalytic cavity is large enough to accommodate 5mC and its derivatives for further oxidation. With the exception of the methyl-CpG dinucleotide, only the DNA phosphate groups are involved in the TET2-DNA contacts. Further biochemical analysis supports that TET2 has no sequence selectivity for the DNA sequence besides the CpG dinucleotide (Hu et al. 2013).

4.3 Crystal Structure of the NgTet1-5mC-DNA Complex

TET enzymes are widely distributed across species, including the heterolobosean amoeboflagellate *Naegleria gruberi*. The genome of *Naegleria* encodes eight Tet-like dioxygenases (NgTet1-8). Sequence analysis has indicated that the NgTet enzymes have a DSBH core region and Fe(II)-chelating residues (HxD…H motif) but lack the Cys-rich region. Biochemical analyses have demonstrated that NgTet1 can successively oxidize 5mC to 5hmC, 5fC, and 5caC using DNA substrates in the XpG (X = 5mC, 5hmC, or 5fC) dinucleotide context (Hashimoto et al. 2014). Hashimoto et al. have determined the crystal structure of NgTet1 in complex with a 14-bp methylated DNA at 2.9 Å resolution (Hashimoto et al. 2014).

NgTet1 has a DSBH core formed by two β-sheets, with the eight-stranded β-sheet stabilized by five α-helices (Fig. 3c). The DNA binds to NgTet1 on the basic surface and the flipped 5mC inserts into the catalytic cavity. The overall structure and the pattern of DNA recognition are similar to that observed in the structure of TET2-DNA complex. The hairpin loop L1 of NgTet1 is equivalent to loop L2 of human TET2, which is important for DNA recognition. NgTet1 lacks the Cys-rich region. As a result, the enzyme lacks the equivalents of loop L1 and the regions involved in the coordination of the three zinc cations in human TET2. Moreover, no insert within the DSBH is present in any of the eight NgTet enzymes. TET enzymes may have gained these additional regions (loop L1 and the insert) for potential regulatory functions during evolution.

4.4 Structural Basis for Substrate Preference in TET-Mediated Oxidation

Previous studies showed that 5hmC is much (~10–100-fold) more abundant than 5fC/5caC (Globisch et al. 2010; Ito et al. 2011; Mellen et al. 2012; Pfaffeneder et al. 2011; Song et al. 2011; Tahiliani et al. 2009). The presence of thymine-DNA glycosylase (TDG) seems not to be predominately responsible for such different abundance of 5hmC and 5fC/5caC, because the depletion of TDG leads to an accumulation of 5fC and 5caC by 2–10-fold, but no apparent changes of the 5hmC and 5mC level in mouse ESCs (Shen et al. 2013). In vitro enzymatic analyses also show that TET enzymes, including human TET1/2, mouse Tet2, and *Naegleria* Tet-like protein,

possess higher activity for DNA substrate containing 5mC than 5hmC/5fC-DNA (Ito et al. 2011; Hashimoto et al. 2014; Hu et al. 2015). These studies suggest that TET enzymes might play a major role in controlling the cellular level of 5mC oxidized derivatives.

Recently, we determined the crystal structures of human TET2 in complex with 5hmC-DNA and 5fC-DNA (Hu et al. 2015) (Fig. 3d). The structural analyses indicate that 5hmC or 5fC is specifically recognized by TET2 in a manner similar to that of 5mC in TET2-5mC-DNA structure (Hu et al. 2013). The cytosine portion of 5mC/5hmC/5fC adopts an almost identical conformation within the catalytic cavity in the three structures (Hu et al. 2013, 2015). The major difference between 5hmC and 5fC is that the hydroxyl group of 5hmC and carbonyl group of 5fC face toward opposite directions, because the hydroxymethyl group of 5hmC and the formyl group of 5fC form hydrogen bonds with 1-carboxylate of NOG and N4 exocyclic nitrogen of cytosine, respectively. The hydrogen bonds prevent the C-C bond between carbon 5 of cytosine and the methyl group (5hmC or 5fC) from free rotation. Therefore, the hydroxymethyl group of 5hmC and the formyl group of 5fC adopt restrained conformations within the catalytic cavity, whereas the methyl group of 5mC is not restrained, because no hydrogen bond is formed. Further biochemical analyses and molecular dynamic simulations suggest that such a restrained conformation may prevent the hydrogen(s) of 5hmC/5fC from adopting an orientation favorable for hydrogen abstraction during catalysis leading to low catalytic efficiency. The residues for catalytic cavity formation are highly conserved, which suggests a general mechanism for TET enzymes. Thus, the substrate preference of TET enzyme results from the type of 5mC derivative groups, and TET enzymes are less reactive toward 5hmC, further supporting that 5hmC may serve as a potentially epigenetic mark for regulatory functions (Hu et al. 2015). Genome-wide analyses have shown that 5fC and 5caC mainly occur at specific regions (Song et al. 2013; Shen et al. 2013), suggesting that TET might be recruited to these regions in higher concentration or be more active through mechanisms yet to be discovered.

5 Regulation of TET Enzymes

5.1 Inhibitors

As a co-substrate, α-KG directly binds to TET enzymes and is converted into succinate and carbon dioxide during each cycle of oxidation. Succinate and α-KG bind to the catalytic cavity of TET enzymes in a similar manner. In various tumors, the pathological accumulation of natural metabolites (succinate and fumarate) and oncometabolites (2-hydroxyglutarate, 2HG) has been observed, which are structurally similar to α-KG and therefore lead to competitive inhibition of TET enzymes (Fig. 2).

Isocitrate dehydrogenases (IDH) are important metabolic enzymes involved in the tricarboxylic acid (TCA) cycle through converting isocitrate to α-KG. The *IDH1/2* genes are frequently mutated in de novo AML (Mardis et al. 2009),

melanoma (Shibata et al. 2011), glioma (Parsons et al. 2008), and thyroid carcinomas (Hemerly et al. 2010). The *IDH1/2* mutations result in a gain of enzymatic activity for the production and accumulation of the oncometabolite R-2-hydroxyglutarate (R-2HG) (Figueroa et al. 2010). The resultant α-KG analog functions as a competitor to inhibit the activities of various α-KG/Fe(II)-dependent dioxygenases, including JmjC-containing histone demethylases and TET enzymes (Xu et al. 2011a; Ye et al. 2013). Loss-of-function mutations of two other key enzymes in the TCA cycle [fumarate hydratase (FH) and succinate dehydrogenase (SDH)] have also been observed in various tumors. These mutations lead to the accumulation of their substrates (fumarate and succinate), which results in similar effects as IDH1/2 mutations (Xiao et al. 2012).

NOG (N-oxalylglycine) is an inactive analog of α-KG and binds to α-KG/Fe(II)-dependent dioxygenases in a manner similar to that of α-KG (Cloos et al. 2006; Hamada et al. 2009). However, it is unable to undergo decarboxylation in the catalytic center and thus is commonly used as an inhibitor to block the enzymatic activity of these enzymes for in vitro biochemical and structural studies. However, NOG inhibits TET enzymes and all other α-KG/Fe(II)-dependent dioxygenases and is not an ideal inhibitor for in vivo studies. Inhibitors with high selectivity (only targeting TET enzymes) would advance the functional studies of TET enzymes under biological and pathological conditions, such as the early stage of embryo development.

5.2 Activators

Vitamin C (also known as L-ascorbic acid) is a dietary nutrient that is critical for mammals. It functions as an antioxidant and plays an important role in biosynthesis of collagen, catecholamine, and carnitine (Englard and Seifter 1986) and facilitates the generation of iPSCs (Esteban et al. 2010). Vitamin C has been reported to enhance the enzymatic activities of a number of α-KG/Fe(II)-dependent dioxygenases, including prolyl 4-hydroxylase (P4H) (Myllyla et al. 1978) and JmjC-containing histone demethylases (Wang et al. 2011). The vitamin C-induced enhancement of TET activity has been observed in vitro and under various physiological conditions, including mouse ESCs and MEFs (Yin et al. 2013; Minor et al. 2013; Blaschke et al. 2013; Chen et al. 2013a) (Fig. 2). Therefore, vitamin C may contribute to epigenetic remodeling through regulating JmjC-containing histone demethylases and TET enzymes (Young et al. 2015). Vitamin C is generally believed to function as an antioxidant to prevent the oxidation of Fe(II). However, the activity enhancement of α-KG/Fe(II)-dependent dioxygenases does not occur when vitamin C is replaced by other reducing agents and antioxidants, such as DTT, glutathione, and L-cysteine (Blaschke et al. 2013; Yin et al. 2013), indicating a specific effect. Vitamin C appears to bind to the C-terminal catalytic domain of TET enzymes (Yin et al. 2013), whereas the underlying mechanism for the activity enhancement needs further investigation.

Adenosine triphosphate (ATP) has been found to enhance the in vitro activity of TET enzymes (He et al. 2011) (Fig. 2). Although the cellular concentration of ATP is high enough for the enhancement, the physiological relevance of this finding

remains unknown. Because the expression and activity of TET enzymes are impaired in various tumors, it is of interest to test whether vitamin C, ATP, or other TET activators yet to be discovered could be used for cancer treatment through enhancing the activity of TET enzymes.

5.3 Interacting Proteins

O-linked β-N-acetylglucosamine transferase (OGT) is an enzyme that transfers O-GlcNac moiety to the hydroxyl groups of threonine and serine residues of various protein substrates for specific regulations (Hanover et al. 2012). OGT directly interacts with the DSBH domain of TET enzymes (primarily Tet2 and Tet3) (Chen et al. 2013b; Deplus et al. 2013; Vella et al. 2013). However, neither the interaction nor the O-GlcNAcylation of TET regulates the activity of TET enzymes. Genome-wide analyses have indicated that TET enzymes recruit OGT to CpG-rich promoters. The depletion of TET enzymes impairs the chromatin association of OGT and OGT-mediated O-GlcNAcylation of histone H2B (Chen et al. 2013b) and HCF1 (host cell factor 1) (Deplus et al. 2013).

Tet1 binds to and co-localize with the SIN3A co-repressor complex. Tet1 contributes to the genomic targeting of SIN3A, while SIN3A does not affect Tet1 binding to the target genes (Williams et al. 2011). Recent studies have indicated that WT1, encoded by Wilms' tumor gene *WT1*, interacts with and recruits TET2 to its target genes and regulates their gene expression (Wang et al. 2015b; Rampal et al. 2014). Interestingly, *WT1*, *TET2*, and *IDH1/2* are mutated in a mutually exclusive manner in AML, suggesting that these genes act in a same pathway for the suppression of oncogenesis.

6 Concluding Remarks

It has now been well established that TET enzymes mediate 5mC oxidation in genome and play important roles in DNA demethylation, gene transcription, embryonic development, and oncogenesis. What remains to be addressed is how the activity and genomic localization of TET enzymes are precisely determined and dynamically regulated, especially during developmental and pathological processes. In particular, what are the key factors to allow TET enzymes to exhibit basal activity to generate 5hmC or higher activity to generate 5fC/5caC in specific genomic regions? What are the roles of 5hmC, 5fC, and 5caC? Are there specific readers to interpret these modifications? Moreover, specific TET inhibitors would provide valuable tools to study whether TET enzymes could be potential drug target for therapeutic applications.

Acknowledgments We thank Dr. Guoliang Xu and his lab members for critical reading of the manuscript. This work was supported by grants from the National Natural Science Foundation of China (31425008 and 91419301). We apologize that we could not cite many important papers due to space limitation.

References

Abdel-Wahab O, Mullally A, Hedvat C, Garcia-Manero G, Patel J, Wadleigh M, et al. Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. Blood. 2009;114(1):144–7. doi:10.1182/blood-2009-03-210039.

Arioka Y, Watanabe A, Saito K, Yamada Y. Activation-induced cytidine deaminase alters the subcellular localization of Tet family proteins. PLoS One. 2012;7(9):e45031. doi:10.1371/journal.pone.0045031.

Bird A. DNA methylation patterns and epigenetic memory. Genes Dev. 2002;16(1):6–21. doi:10.1101/gad.947102.

Blaschke K, Ebata KT, Karimi MM, Zepeda-Martinez JA, Goyal P, Mahapatra S, et al. Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. Nature. 2013;500(7461):222–6. doi:10.1038/nature12362.

Branco MR, Ficz G, Reik W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. Nat Rev Genet. 2012;13(1):7–13. doi:10.1038/nrg3080.

Bruniquel D, Schwartz RH. Selective, stable demethylation of the interleukin-2 gene enhances transcription by an active process. Nat Immunol. 2003;4(3):235–40. doi:10.1038/ni887.

Chen CC, Wang KY, Shen CKJ. The Mammalian de Novo DNA methyltransferases DNMT3A and DNMT3B are also DNA 5-hydroxymethylcytosine dehydroxymethylases. J Biol Chem. 2012;287(40):33116–21. doi:10.1074/Jbc.C112.406975.

Chen J, Guo L, Zhang L, Wu H, Yang J, Liu H, et al. Vitamin C modulates TET1 function during somatic cell reprogramming. Nat Genet. 2013a;45(12):1504–9. doi:10.1038/ng.2807.

Chen Q, Chen Y, Bian C, Fujiki R, Yu X. TET2 promotes histone O-GlcNAcylation during gene transcription. Nature. 2013b;493(7433):561–4. doi:10.1038/nature11742.

Cimmino L, Abdel-Wahab O, Levine RL, Aifantis I. TET family proteins and their role in stem cell differentiation and transformation. Cell Stem Cell. 2011;9(3):193–204. doi:10.1016/J.Stem.2011.08.007.

Cliffe LJ, Kieft R, Southern T, Birkeland SR, Marshall M, Sweeney K, et al. JBP1 and JBP2 are two distinct thymidine hydroxylases involved in J biosynthesis in genomic DNA of African trypanosomes. Nucleic Acids Res. 2009;37(5):1452–62. doi:10.1093/nar/gkn1067.

Cloos PA, Christensen J, Agger K, Maiolica A, Rappsilber J, Antal T, et al. The putative oncogene GASC1 demethylates tri- and dimethylated lysine 9 on histone H3. Nature. 2006;442(7100):307–11. doi:10.1038/nature04837.

Costa Y, Ding J, Theunissen TW, Faiola F, Hore TA, Shliaha PV, et al. NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. Nature. 2013;495(7441):370–4. doi:10.1038/nature11925.

Dawlaty MM, Ganz K, Powell BE, Hu YC, Markoulaki S, Cheng AW, et al. Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. Cell Stem Cell. 2011;9(2):166–75. doi:10.1016/J.Stem.2011.07.010.

Dawlaty MM, Breiling A, Le T, Raddatz G, Barrasa MI, Cheng AW, et al. Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. Dev Cell. 2013;24(3):310–23. doi:10.1016/J.Devcel.2012.12.015.

Deplus R, Delatte B, Schwinn MK, Defrance M, Mendez J, Murphy N, et al. TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. EMBO J. 2013;32(5):645–55. doi:10.1038/emboj.2012.357.

Doege CA, Inoue K, Yamashita T, Rhee DB, Travis S, Fujita R, et al. Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. Nature. 2012;488(7413):652–5. doi:10.1038/nature11333.

Englard S, Seifter S. The biochemical functions of ascorbic acid. Annu Rev Nutr. 1986;6:365–406. doi:10.1146/annurev.nu.06.070186.002053.

Esteban MA, Wang T, Qin B, Yang J, Qin D, Cai J, et al. Vitamin C enhances the generation of mouse and human induced pluripotent stem cells. Cell Stem Cell. 2010;6(1):71–9. doi:10.1016/j.stem.2009.12.001.

Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, et al. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature. 2011;473(7347):398–402. doi:10.1038/nature10008.

Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, Shih A, et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. Cancer Cell. 2010;18(6):553–67. doi:10.1016/j.ccr.2010.11.015.

Frauer C, Hoffmann T, Bultmann S, Casa V, Cardoso MC, Antes I, et al. Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. PLoS One. 2011;6(6):e21306. doi:10.1371/journal.pone.0021306.

Fu L, Guerrero CR, Zhong N, Amato NJ, Liu Y, Liu S, et al. Tet-mediated formation of 5-hydroxymethylcytosine in RNA. J Am Chem Soc. 2014;136(33):11582–5. doi:10.1021/ja505305z.

Fu Y, Luo GZ, Chen K, Deng X, Yu M, Han D, et al. N6-methyldeoxyadenosine marks active transcription start sites in chlamydomonas. Cell. 2015;161(4):879–92. doi:10.1016/j.cell.2015.04.010.

Gao Y, Chen J, Li K, Wu T, Huang B, Liu W, et al. Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. Cell Stem Cell. 2013;12(4):453–69. doi:10.1016/j.stem.2013.02.005.

Globisch D, Munzel M, Muller M, Michalakis S, Wagner M, Koch S, et al. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PLoS One. 2010;5(12):e15367. doi:10.1371/journal.pone.0015367.

Greer EL, Blanco MA, Gu L, Sendinc E, Liu J, Aristizabal-Corrales D, et al. DNA Methylation on N6-Adenine in C. elegans. Cell. 2015;161(4):868–78. doi:10.1016/j.cell.2015.04.005.

Gu TP, Guo F, Yang H, Wu HP, Xu GF, Liu W, et al. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature. 2011;477(7366):606–10. doi:10.1038/nature10443.

Guo JU, Su Y, Zhong C, Ming GL, Song H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell. 2011;145(3):423–34. doi:10.1016/j.cell.2011.03.022.

Guo F, Li X, Liang D, Li T, Zhu P, Guo H, et al. Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. Cell Stem Cell. 2014;15(4):447–58. doi:10.1016/j.stem.2014.08.003.

Hackett JA, Sengupta R, Zylicz JJ, Murakami K, Lee C, Down TA, et al. Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. Science. 2013;339(6118):448–52. doi:10.1126/science.1229277.

Haffner MC, Chaux A, Meeker AK, Esopi DM, Gerber J, Pellakuru LG, et al. Global 5-hydroxymethylcytosine content is significantly reduced in tissue stem/progenitor cell compartments and in human cancers. Oncotarget. 2011;2(8):627–37.

Hajkova P, Erhardt S, Lane N, Haaf T, El-Maarri O, Reik W, et al. Epigenetic reprogramming in mouse primordial germ cells. Mech Dev. 2002;117(1–2):15–23.

Hamada S, Kim TD, Suzuki T, Itoh Y, Tsumoto H, Nakagawa H, et al. Synthesis and activity of N-oxalylglycine and its derivatives as Jumonji C-domain-containing histone lysine demethylase inhibitors. Bioorg Med Chem Lett. 2009;19(10):2852–5. doi:10.1016/j.bmcl.2009.03.098.

Hanover JA, Krause MW, Love DC. Bittersweet memories: linking metabolism to epigenetics through O-GlcNAcylation. Nat Rev Mol Cell Biol. 2012;13(5):312–21. doi:10.1038/nrm3334.

Hashimoto H, Liu Y, Upadhyay AK, Chang Y, Howerton SB, Vertino PM, et al. Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Res. 2012;40(11):4841–9. doi:10.1093/nar/gks155.

Hashimoto H, Pais JE, Zhang X, Saleh L, Fu ZQ, Dai N, et al. Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. Nature. 2014;506(7488):391–5. doi:10.1038/nature12905.

He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science. 2011;333(6047):1303–7. doi:10.1126/science.1210944.

Hemerly JP, Bastos AU, Cerutti JM. Identification of several novel non-p.R132 IDH1 variants in thyroid carcinomas. Eur J Endocrinol. 2010;163(5):747–55. doi:10.1530/Eje-10-0473.
Hoffart LM, Barr EW, Guyer RB, Bollinger JM, Krebs C. Direct spectroscopic detection of a CH-cleaving high-spin Fe (IV) complex in a prolyl-4-hydroxylase. Proc Natl Acad Sci. 2006;103(40):14738–43.
Hoon DS, Spugnardi M, Kuo C, Huang SK, Morton DL, Taback B. Profiling epigenetic inactivation of tumor suppressor genes in tumors and plasma from cutaneous melanoma patients. Oncogene. 2004;23(22):4014–22. doi:10.1038/sj.onc.1207505.
Hu L, Li Z, Cheng J, Rao Q, Gong W, Liu M, et al. Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. Cell. 2013;155(7):1545–55. doi:10.1016/j.cell.2013.11.020.
Hu X, Zhang L, Mao SQ, Li Z, Chen J, Zhang RR, et al. Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. Cell Stem Cell. 2014;14(4):512–22. doi:10.1016/j.stem.2014.01.001.
Hu L, Lu J, Cheng J, Rao Q, Li Z, Hou H, et al. Structural insight into substrate preference for TET-mediated oxidation. Nature. 2015;527(7576):118–22. doi:10.1038/nature15713.
Inoue A, Zhang Y. Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. Science. 2011;334(6053):194. doi:10.1126/science.1212483.
Inoue A, Shen L, Dai Q, He C, Zhang Y. Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. Cell Res. 2011;21(12):1670–6. doi:10.1038/cr.2011.189.
Iqbal K, Jin SG, Pfeifer GP, Szabo PE. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. Proc Natl Acad Sci U S A. 2011;108(9):3642–7. doi:10.1073/Pnas.1014033108.
Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature. 2010;466(7310):1129–33. doi:10.1038/nature09303.
Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science. 2011;333(6047):1300–3. doi:10.1126/Science.1210597.
Iyer LM, Tahiliani M, Rao A, Aravind L. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. Cell Cycle. 2009;8(11):1698–710.
Iyer LM, Zhang D, Burroughs AM, Aravind L. Computational identification of novel biochemical systems involved in oxidation, glycosylation and other complex modifications of bases in DNA. Nucleic Acids Res. 2013;41(16):7635–55. doi:10.1093/nar/gkt573.
Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet. 2003;33(Suppl):245–54. doi:10.1038/ng1089.
Kaas GA, Zhong C, Eason DE, Ross DL, Vachhani RV, Ming GL, et al. TET1 controls CNS 5-methylcytosine hydroxylation, active DNA demethylation, gene transcription, and memory formation. Neuron. 2013;79(6):1086–93. doi:10.1016/j.neuron.2013.08.032.
Kangaspeska S, Stride B, Metivier R, Polycarpou-Schwarz M, Ibberson D, Carmouche RP, et al. Transient cyclical methylation of promoter DNA. Nature. 2008;452(7183):112–5. doi:10.1038/nature06640.
Ko M, Huang Y, Jankowska AM, Pape UJ, Tahiliani M, Bandukwala HS, et al. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature. 2010;468(7325):839–43. doi:10.1038/nature09586.
Ko M, An J, Bandukwala HS, Chavez L, Aijo T, Pastor WA, et al. Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. Nature. 2013;497(7447):122–6. doi:10.1038/nature12052.
Koh KP, Yabuuchi A, Rao S, Huang Y, Cunniff K, Nardone J, et al. Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. Cell Stem Cell. 2011;8(2):200–13. doi:10.1016/j.stem.2011.01.008.

Krebs C, Galonic Fujimori D, Walsh CT, Bollinger Jr JM. Non-heme Fe(IV)-oxo intermediates. Acc Chem Res. 2007;40(7):484–92. doi:10.1021/ar700066p.

Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science. 2009;324(5929):929–30. doi:10.1126/science.1169786.

Kurimoto K, Yabuta Y, Ohinata Y, Shigeta M, Yamanaka K, Saitou M. Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. Genes Dev. 2008;22(12):1617–35. doi:10.1101/gad.1649908.

Langemeijer SM, Kuiper RP, Berends M, Knops R, Aslanyan MG, Massop M, et al. Acquired mutations in TET2 are common in myelodysplastic syndromes. Nat Genet. 2009;41(7):838–42. doi:10.1038/ng.391.

Lee J, Inoue K, Ono R, Ogonuki N, Kohda T, Kaneko-Ishino T, et al. Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. Development. 2002;129(8):1807–17.

Lian CG, Xu Y, Ceol C, Wu F, Larson A, Dresser K, et al. Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. Cell. 2012;150(6):1135–46. doi:10.1016/j.cell.2012.07.033.

Liu CK, Hsu CA, Abbott MT. Catalysis of three sequential dioxygenase reactions by thymine 7-hydroxylase. Arch Biochem Biophys. 1973;159(1):180–7.

Liu S, Ren S, Howell P, Fodstad O, Riker AI. Identification of novel epigenetically modified genes in human melanoma via promoter methylation gene profiling. Pigment Cell Melanoma Res. 2008;21(5):545–58. doi:10.1111/j.1755-148X.2008.00484.x.

Liutkeviciute Z, Lukinavicius G, Masevicius V, Daujotyte D, Klimasauskas S. Cytosine-5-methyltransferases add aldehydes to DNA. Nat Chem Biol. 2009;5(6):400–2. doi:10.1038/nchembio.172.

Liutkeviciute Z, Kriukiene E, Licyte J, Rudyte M, Urbanaviciute G, Klimasauskas S. Direct decarboxylation of 5-carboxylcytosine by DNA C5-methyltransferases. J Am Chem Soc. 2014;136(16):5884–7. doi:10.1021/ja5019223.

Loenarz C, Schofield CJ. Expanding chemical biology of 2-oxoglutarate oxygenases. Nat Chem Biol. 2008;4(3):152–6. doi:10.1038/nchembio0308-152.

Long HK, Blackledge NP, Klose RJ. ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. Biochem Soc Trans. 2013;41(3):727–40. doi:10.1042/BST20130028.

Lu X, Han D, Zhao BS, Song CX, Zhang LS, Dore LC, et al. Base-resolution maps of 5-formylcytosine and 5-carboxylcytosine reveal genome-wide DNA demethylation dynamics. Cell Res. 2015;25(3):386–9. doi:10.1038/cr.2015.5.

Maiti A, Drohat AC. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine potential implications for active demethylation of CpG sites. J Biol Chem. 2011;286(41):35334–8. doi:10.1074/Jbc.C111.284620.

Mardis ER, Ding L, Dooling DJ, Larson DE, McLellan MD, Chen K, et al. Recurring mutations found by sequencing an acute myeloid leukemia genome. N Engl J Med. 2009;361(11):1058–66. doi:10.1056/Nejmoa0903840.

Martinowich K, Hattori D, Wu H, Fouse S, He F, Hu Y, et al. DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. Science. 2003;302(5646):890–3. doi:10.1126/science.1090842.

Mayer W, Niveleau A, Walter J, Fundele R, Haaf T. Embryogenesis - demethylation of the zygotic paternal genome. Nature. 2000;403(6769):501–2. doi:10.1038/35000656.

McDonough MA, Loenarz C, Chowdhury R, Clifton IJ, Schofield CJ. Structural studies on human 2-oxoglutarate dependent oxygenases. Curr Opin Struct Biol. 2010;20(6):659–72. doi:10.1016/j.sbi.2010.08.006.

Mellen M, Ayata P, Dewell S, Kriaucionis S, Heintz N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. Cell. 2012;151(7):1417–30. doi:10.1016/j.cell.2012.11.022.

Metivier R, Gallais R, Tiffoche C, Le Peron C, Jurkowska RZ, Carmouche RP, et al. Cyclical DNA methylation of a transcriptionally active promoter. Nature. 2008;452(7183):45–50. doi:10.1038/nature06544.

Minor EA, Court BL, Young JI, Wang G. Ascorbate induces ten-eleven translocation (Tet) methylcytosine dioxygenase-mediated generation of 5-hydroxymethylcytosine. J Biol Chem. 2013;288(19):13669–74. doi:10.1074/jbc.C113.464800.

Moran-Crusio K, Reavie L, Shih A, Abdel-Wahab O, Ndiaye-Lobry D, Lobry C, et al. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. Cancer Cell. 2011;20(1):11–24. doi:10.1016/j.ccr.2011.06.001.

Muller I, Stuckl C, Wakeley J, Kertesz M, Uson I. Succinate complex crystal structures of the alpha-ketoglutarate-dependent dioxygenase AtsK: steric aspects of enzyme self-hydroxylation. J Biol Chem. 2005;280(7):5716–23. doi:10.1074/jbc.M410840200.

Munzel M, Globisch D, Bruckl T, Wagner M, Welzmiller V, Michalakis S, et al. Quantification of the sixth DNA base hydroxymethylcytosine in the brain. Angew Chem Int Ed Engl. 2010;49(31):5375–7. doi:10.1002/anie.201002033.

Myllyla R, Kuutti-Savolainen ER, Kivirikko KI. The role of ascorbate in the prolyl hydroxylase reaction. Biochem Biophys Res Commun. 1978;83(2):441–8.

Nabel CS, Jia H, Ye Y, Shen L, Goldschmidt HL, Stivers JT, et al. AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. Nat Chem Biol. 2012;8(9):751–8. doi:10.1038/nchembio.1042.

Nakamura T, Arai Y, Umehara H, Masuhara M, Kimura T, Taniguchi H, et al. PGC7/Stella protects against DNA demethylation in early embryogenesis. Nat Cell Biol. 2007;9(1):64–71. doi:10.1038/ncb1519.

Neidigh JW, Darwanto A, Williams AA, Wall NR, Sowers LC. Cloning and characterization of Rhodotorula glutinis thymine hydroxylase. Chem Res Toxicol. 2009;22(5):885–93. doi:10.1021/tx8004482.

Ono R, Taki T, Taketani T, Taniwaki M, Kobayashi H, Hayashi Y. LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). Cancer Res. 2002;62(14):4075–80.

Oswald J, Engemann S, Lane N, Mayer W, Olek A, Fundele R, et al. Active demethylation of the paternal genome in the mouse zygote. Curr Biol. 2000;10(8):475–8.

Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, et al. An integrated genomic analysis of human glioblastoma multiforme. Science. 2008;321(5897):1807–12. doi:10.1126/science.1164382.

Pastor WA, Aravind L, Rao A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. Nat Rev Mol Cell Biol. 2013;14(6):341–56. doi:10.1038/nrm3589.

Pfaffeneder T, Hackner B, Truss M, Munzel M, Muller M, Deiml CA, et al. The discovery of 5-formylcytosine in embryonic stem cell DNA. Angew Chem Int Ed Engl. 2011;50(31):7008–12. doi:10.1002/anie.201103899.

Price JC, Barr EW, Glass TE, Krebs C, Bollinger Jr JM. Evidence for hydrogen abstraction from C1 of taurine by the high-spin Fe(IV) intermediate detected during oxygen activation by taurine:alpha-ketoglutarate dioxygenase (TauD). J Am Chem Soc. 2003;125(43):13008–9. doi:10.1021/ja037400h.

Quivoron C, Couronne L, Della Valle V, Lopez CK, Plo I, Wagner-Ballon O, et al. TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. Cancer Cell. 2011;20(1):25–38. doi:10.1016/j.ccr.2011.06.003.

Rampal R, Alkalin A, Madzo J, Vasanthakumar A, Pronier E, Patel J, et al. DNA hydroxymethylation profiling reveals that WT1 mutations result in loss of TET2 function in acute myeloid leukemia. Cell Rep. 2014;9(5):1841–55. doi:10.1016/j.celrep.2014.11.004.

Rangam G, Schmitz KM, Cobb AJ, Petersen-Mahrt SK. AID enzymatic activity is inversely proportional to the size of cytosine C5 orbital cloud. PLoS One. 2012;7(8):e43279. doi:10.1371/journal.pone.0043279.

Rudenko A, Dawlaty MM, Seo J, Cheng AW, Meng J, Le T, et al. Tet1 is critical for neuronal activity-regulated gene expression and memory extinction. Neuron. 2013;79(6):1109–22. doi:10.1016/J.Neuron.2013.08.003.

Ruzov A, Tsenkina Y, Serio A, Dudnakova T, Fletcher J, Bai Y, et al. Lineage-specific distribution of high levels of genomic 5-hydroxymethylcytosine in mammalian development. Cell Res. 2011;21(9):1332–42. doi:10.1038/cr.2011.113.

Ryle MJ, Padmakumar R, Hausinger RP. Stopped-flow kinetic analysis of Escherichia coli taurine/alpha-ketoglutarate dioxygenase: interactions with alpha-ketoglutarate, taurine, and oxygen. Biochemistry. 1999;38(46):15278–86.

Saitou M, Kagiwada S, Kurimoto K. Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. Development. 2012;139(1):15–31. doi:10.1242/dev.050849.

Schiesser S, Hackner B, Pfaffeneder T, Muller M, Hagemeier C, Truss M, et al. Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. Angew Chem Int Ed Engl. 2012;51(26):6516–20. doi:10.1002/anie.201202583.

Seisenberger S, Andrews S, Krueger F, Arand J, Walter J, Santos F, et al. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol Cell. 2012;48(6):849–62. doi:10.1016/j.molcel.2012.11.001.

Shen L, Kondo Y, Guo Y, Zhang J, Zhang L, Ahmed S, et al. Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. PLoS Genet. 2007;3(10):2023–36. doi:10.1371/journal.pgen.0030181.

Shen L, Wu H, Diep D, Yamaguchi S, D'Alessio AC, Fung HL, et al. Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. Cell. 2013;153(3):692–706. doi:10.1016/j.cell.2013.04.002.

Shen L, Inoue A, He J, Liu YT, Lu FL, Zhang Y. Tet3 and DNA replication mediate demethylation of both the maternal and paternal genomes in mouse zygotes. Cell Stem Cell. 2014a;15(4):459–70. doi:10.1016/J.Stem.2014.09.002.

Shen L, Song CX, He C, Zhang Y. Mechanism and function of oxidative reversal of DNA and RNA methylation. Annu Rev Biochem. 2014b;83:585–614. doi:10.1146/annurev-biochem-060713-035513.

Shibata T, Kokubu A, Miyamoto M, Sasajima Y, Yamazaki N. Mutant IDH1 confers an in vivo growth in a melanoma cell line with BRAF mutation. Am J Pathol. 2011;178(3):1395–402. doi:10.1016/j.ajpath.2010.12.011.

Smiley JA, Kundracik M, Landfried DA, Barnes Sr VR, Axhemi AA. Genes of the thymidine salvage pathway thymine 7 hydroxylase from a Rhodotorula glutinis cDNA library and iso-orotate decarboxylase from Neurospora crassa. Biochim Biophys Acta. 2005;1723(1–3):256–64. doi:10.1016/j.bbagen.2005.02.001.

Smith ZD, Meissner A. DNA methylation: roles in mammalian development. Nat Rev Genet. 2013;14(3):204–20. doi:10.1038/nrg3354.

Song CX, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, et al. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat Biotechnol. 2011;29(1):68–72. doi:10.1038/nbt.1732.

Song CX, Szulwach KE, Dai Q, Fu Y, Mao SQ, Lin L, et al. Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. Cell. 2013;153(3):678–91. doi:10.1016/j.cell.2013.04.001.

Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. Nucleic Acids Res. 2010;38(19):e181. doi:10.1093/nar/gkq684.

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2009;324(5929):930–5. doi:10.1126/science.1170116.

Tan L, Shi YG. Tet family proteins and 5-hydroxymethylcytosine in development and disease. Development. 2012;139(11):1895–902. doi:10.1242/dev.070771.

Tefferi A, Levine RL, Lim KH, Abdel-Wahab O, Lasho TL, Patel J, et al. Frequent TET2 mutations in systemic mastocytosis: clinical, KITD816V and FIP1L1-PDGFRA correlates. Leukemia. 2009;23(5):900–4. doi:10.1038/leu.2009.37.

Tsai CL, Tainer JA. Probing DNA by 2-OG-dependent dioxygenase. Cell. 2013;155(7):1448–50. doi:10.1016/j.cell.2013.12.002.

Valegard K, van Scheltinga AC T, Dubus A, Ranghino G, Oster LM, Hajdu J, et al. The structural basis of cephalosporin formation in a mononuclear ferrous enzyme. Nat Struct Mol Biol. 2004;11(1):95–101. doi:10.1038/nsmb712.

Valinluck V, Sowers LC. Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. Cancer Res. 2007;67(3):946–50. doi:10.1158/0008-5772.Can-06-3123.

Valinluck V, Tsai HH, Rogstad DK, Burdzy A, Bird A, Sowers LC. Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). Nucleic Acids Res. 2004;32(14):4100–8. doi:10.1093/nar/gkh739.

Vella P, Scelfo A, Jammula S, Chiacchiera F, Williams K, Cuomo A, et al. Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. Mol Cell. 2013;49(4):645–56. doi:10.1016/j.molcel.2012.12.019.

Wang T, Chen K, Zeng X, Yang J, Wu Y, Shi X, et al. The histone demethylases Jhdm1a/1b enhance somatic cell reprogramming in a vitamin-C-dependent manner. Cell Stem Cell. 2011;9(6):575–87. doi:10.1016/j.stem.2011.10.005.

Wang L, Zhang J, Duan J, Gao X, Zhu W, Lu X, et al. Programming and inheritance of parental DNA methylomes in mammals. Cell. 2014;157(4):979–91. doi:10.1016/j.cell.2014.04.017.

Wang L, Zhou Y, Xu L, Xiao R, Lu X, Chen L, et al. Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex. Nature. 2015a. doi:10.1038/nature14482.

Wang Y, Xiao M, Chen X, Chen L, Xu Y, Lv L, et al. WT1 recruits TET2 to regulate its target gene expression and suppress leukemia cell proliferation. Mol Cell. 2015b;57(4):662–73. doi:10.1016/j.molcel.2014.12.023.

Williams K, Christensen J, Pedersen MT, Johansen JV, Cloos PA, Rappsilber J, et al. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature. 2011;473(7347):343–8. doi:10.1038/nature10066.

Williams K, Christensen J, Helin K. DNA methylation: TET proteins-guardians of CpG islands? EMBO Rep. 2012;13(1):28–35. doi:10.1038/embor.2011.233.

Wossidlo M, Nakamura T, Lepikhov K, Marques CJ, Zakhartchenko V, Boiani M, et al. 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nat Commun. 2011;2:241. doi:10.1038/Ncomms1240. Artn 241.

Wu SC, Zhang Y. Active DNA demethylation: many roads lead to Rome. Nat Rev Mol Cell Biol. 2010;11(9):607–20. doi:10.1038/nrm2950.

Wu H, Zhang Y. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. Genes Dev. 2011;25(23):2436–52. doi:10.1101/gad.179184.111.

Wu H, Zhang Y. Reversing DNA methylation: mechanisms, genomics, and biological functions. Cell. 2014;156(1–2):45–68. doi:10.1016/j.cell.2013.12.019.

Wu H, D'Alessio AC, Ito S, Xia K, Wang Z, Cui K, et al. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature. 2011;473(7347):389–93. doi:10.1038/nature09934.

Xiao M, Yang H, Xu W, Ma S, Lin H, Zhu H, et al. Inhibition of alpha-KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors. Genes Dev. 2012;26(12):1326–38. doi:10.1101/gad.191056.112.

Xie W, Barr CL, Kim A, Yue F, Lee AY, Eubanks J, et al. Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. Cell. 2012;148(4):816–31. doi:10.1016/j.cell.2011.12.035.

Xu GL, Walsh CP. Enzymatic DNA oxidation: mechanisms and biological significance. BMB Rep. 2014;47(11):609–18.

Xu W, Yang H, Liu Y, Yang Y, Wang P, Kim SH, et al. Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. Cancer Cell. 2011a;19(1):17–30. doi:10.1016/j.ccr.2010.12.014.

Xu Y, Wu F, Tan L, Kong L, Xiong L, Deng J, et al. Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. Mol Cell. 2011b;42(4):451–64. doi:10.1016/j.molcel.2011.04.005.

Xu Y, Xu C, Kato A, Tempel W, Abreu JG, Bian C, et al. Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development. Cell. 2012;151(6):1200–13. doi:10.1016/j.cell.2012.11.014.

Xu S, Li W, Zhu J, Wang R, Li Z, Xu GL, et al. Crystal structures of isoorotate decarboxylases reveal a novel catalytic mechanism of 5-carboxyl-uracil decarboxylation and shed light on the search for DNA decarboxylase. Cell Res. 2013;23(11):1296–309. doi:10.1038/cr.2013.107.

Yamaguchi S, Hong K, Liu R, Shen L, Inoue A, Diep D, et al. Tet1 controls meiosis by regulating meiotic gene expression. Nature. 2012;492(7429):443–7. doi:10.1038/nature11709.

Yamaguchi S, Shen L, Liu Y, Sendler D, Zhang Y. Role of Tet1 in erasure of genomic imprinting. Nature. 2013;504(7480):460–4. doi:10.1038/nature12805.

Yamazaki Y, Mann MRW, Lee SS, Marh J, McCarrey JR, Yanagimachi R, et al. Reprogramming of primordial germ cells begins before migration into the genital ridge, making these cells inadequate donors for reproductive cloning. Proc Natl Acad Sci U S A. 2003;100(21):12207–12. doi:10.1073/Pnas.2035119100.

Yang CG, Yi C, Duguid EM, Sullivan CT, Jian X, Rice PA, et al. Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. Nature. 2008;452(7190):961–5. doi:10.1038/nature06889.

Yang CG, Garcia K, He C. Damage detection and base flipping in direct DNA alkylation repair. Chembiochem Eur J Chem Biol. 2009;10(3):417–23. doi:10.1002/cbic.200800580.

Yang H, Liu Y, Bai F, Zhang JY, Ma SH, Liu J, et al. Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. Oncogene. 2013;32(5):663–9. doi:10.1038/onc.2012.67.

Ye D, Ma S, Xiong Y, Guan KL. R-2-hydroxyglutarate as the key effector of IDH mutations promoting oncogenesis. Cancer Cell. 2013;23(3):274–6. doi:10.1016/j.ccr.2013.03.005.

Yildirim O, Li R, Hung JH, Chen PB, Dong X, Ee LS, et al. Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. Cell. 2011;147(7):1498–510. doi:10.1016/j.cell.2011.11.054.

Yin R, Mao SQ, Zhao B, Chong Z, Yang Y, Zhao C, et al. Ascorbic acid enhances Tet-mediated 5-methylcytosine oxidation and promotes DNA demethylation in mammals. J Am Chem Soc. 2013;135(28):10396–403. doi:10.1021/ja4028346.

Young JI, Zuchner S, Wang G. Regulation of the Epigenome by Vitamin C. Annu Rev Nutr. 2015;35:545–64. doi:10.1146/annurev-nutr-071714-034228.

Yu Z, Genest PA, ter Riet B, Sweeney K, DiPaolo C, Kieft R, et al. The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. Nucleic Acids Res. 2007;35(7):2107–15. doi:10.1093/Nar/Gkm049.

Yu M, Hon GC, Szulwach KE, Song CX, Zhang L, Kim A, et al. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. Cell. 2012;149(6):1368–80. doi:10.1016/j.cell.2012.04.027.

Zhang L, Lu X, Lu J, Liang H, Dai Q, Xu GL, et al. Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. Nat Chem Biol. 2012;8(4):328–30. doi:10.1038/nchembio.914.

Zhang G, Huang H, Liu D, Cheng Y, Liu X, Zhang W, et al. N6-methyladenine DNA modification in Drosophila. Cell. 2015;161(4):893–906. doi:10.1016/j.cell.2015.04.018.

Zhou T, Xiong J, Wang M, Yang N, Wong J, Zhu B, et al. Structural basis for hydroxymethylcytosine recognition by the SRA domain of UHRF2. Mol Cell. 2014;54(5):879–86. doi:10.1016/j.molcel.2014.04.003.

Proteins That Read DNA Methylation

Takashi Shimbo and Paul A. Wade

Abstract

Covalent modification of DNA via deposition of a methyl group at the 5′ position on cytosine residues alters the chemical groups available for interaction in the major groove of DNA. The information content inherent in this modification alters the affinity and the specificity of DNA binding; some proteins favor interaction with methylated DNA, and others disfavor it. Molecular recognition of cytosine methylation by proteins often initiates sequential regulatory events which impact gene expression and chromatin structure. The known methyl-DNA-binding proteins have unique domains responsible for DNA methylation recognition: (1) the methyl-CpG-binding domain (MBD), (2) the C2H2 zinc finger domain, and (3) the SET- and RING finger-associated (SRA) domain. Structural analyses have revealed that each domain has a characteristic methylated DNA-binding pattern, and this difference in the recognition mechanism renders the DNA methylation mark able to transmit complicated biological information. Recent genetic and genomic studies have revealed novel functions of methyl-DNA-binding proteins. These emerging data have also provided glimpses into how methyl-DNA-binding proteins possess unique features and, presumably, functions. In this review, we summarize structural and biochemical analyses elucidating the mechanism for recognition of DNA methylation and correlate this information with emerging genomic and functional data.

T. Shimbo • P.A. Wade (✉)
Epigenetics and Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, Durham, NC, USA
e-mail: wadep2@niehs.nih.gov

© Springer International Publishing Switzerland 2016

A. Jeltsch, R.Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*, Advances in Experimental Medicine and Biology 945,
DOI 10.1007/978-3-319-43624-1_13

1 Introduction

DNA methylation serves as a fundamental component of epigenetic regulation; dysregulation of DNA methylation impacts multiple biological processes, including tumorigenesis (Schübeler 2015). In mammals, most DNA methylation occurs in the context of the simple palindrome 5′-CG-3′, the CpG dinucleotide. In general, 70–80 % of the CpGs in mammalian genomes are methylated (Bird 2002). Importantly, although DNA methylation has historically been depicted as a relatively static modification, recent studies have revealed that the methyl group on cytosine can be further modified by oxidation; Fe (II)- and α-ketoglutarate-dependent oxidation mediated by ten-eleven translocation (TET) dioxygenases converts 5-methylcytosine (5mC) into 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) (Tahiliani et al. 2009; Kriaucionis and Tahiliani 2014; Kohli and Zhang 2013).

The "reader proteins," referred to as methylcytosine-binding proteins (MBPs), specifically recognize DNA methylation marks and initiate signaling pathways. MBPs often interact with other proteins and serve as hubs to organize effector proteins to particular loci. It is the particular collection of the effector proteins associated with each MBP protein, and not the act of binding methylated CpG per se that typically elicits a downstream transcriptional effect. The MBP proteins can be classified using structural information into three major families, each characterized by the presence of a critical recognition domain: the methyl-CpG-binding domain (MBD), the C2H2 zinc finger domain, and the SET- and RING finger-associated (SRA) domain. Each MBP has unique features, including DNA-binding preferences, expression patterns, or protein-protein interaction partners, and has critical roles in various biological contexts. The domain architecture of each protein family is unique, presenting opportunities for a single modification, CpG methylation, to nucleate different effectors (Fig. 1).

Many proteins are effectively prevented from high-affinity interaction with their cognate DNA recognition sequences when those loci are methylated. Examples include myc, CREB, and members of the E2F family (Tate and Bird 1993). Recently, Schübeler and colleagues have elegantly clarified the relationship between transcription factor binding site methylation and productive binding for NRF1 in embryonic stem cells, concluding that methylation and productive binding "compete" to establish appropriate regulatory states (Domcke et al. 2015). How the addition of methyl groups impacts DNA-protein transactions in this manner is a fascinating topic beyond the scope of this chapter.

2 The Methyl-CpG-Binding Domain Family

2.1 MeCP2

MeCP2 was the first MBP to be purified biochemically and has its cDNA cloned and sequenced (Meehan et al. 1989, 1992; Lewis et al. 1992). The cDNA initially

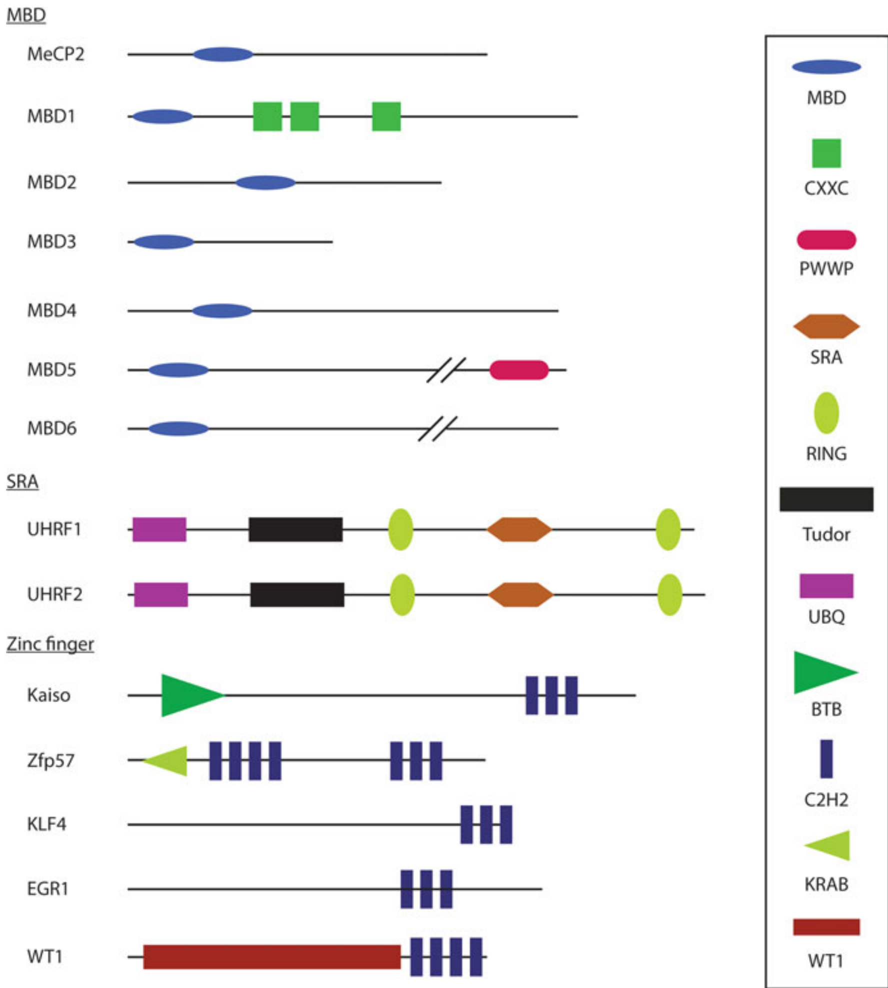


Fig. 1 Methyl-binding proteins and their protein domains. *MBD* Methyl-CpG-binding domain, *CXXC*, CXXC type zinc finger domain, *PWWP* "Pro-Trp-Trp-Pro" domain, *SRA* SET- and RING finger-associated domain, *RING* Really Interesting New Gene finger domain, *Tudor* Tudor domain, *UBQ* ubiquitin-like domain, *BTB* BR-C, ttk, and bab domain, *C2H2* C2H2 type zinc finger domain, *KRAB* Kruppel-associated box domain, *WT1* Wilms' tumor 1 type zinc finger domain

cloned by Bird and colleagues codes for a protein of 492 amino acids that contains an amino terminal MBD domain and a transcriptional repression domain in the carboxyl-terminal region. Somewhat surprisingly, MeCP2 was subsequently found to be homologous to a matrix attachment binding protein from chicken known as ARBP, a protein identified by biochemical means based on its binding to a sequence found in a matrix attachment region (von Kries et al. 1991; Weitzel et al. 1997). The biochemical characteristics initially described for MeCP2 indicated a preference for DNA with a single fully methylated CpG dinucleotide, while the less

well-characterized MeCP1 activity required at least 12 methylated CpG dinucleotides (Meehan et al. 1989). The region responsible for the binding of methylated CpG sites was subsequently identified and termed the MBD (for methyl-CpG-binding domain); this domain became the archetypal methyl-CpG-binding domain. Subsequent homology searches using the MBD from MeCP2 led to the identification of the remaining MBD family proteins (Nan et al. 1993; Hendrich and Bird 1998). MeCP2 thus represents the founding member of the MBD protein family.

The crystal structure analysis of MeCP2 revealed that MeCP2 recognizes the fully methylated CpG dinucleotide using the 5mC-Arg-Gua triad (Fig. 2); two arginines within the MBD domain (R111 and R133) each bind to a guanine with a hydrogen bond and form van der Waals contacts with methylated cytosines (Ho et al. 2008; Liu et al. 2013b). In addition, a tyrosine residue forms water-mediated hydrogen bonds with one of the two cytosine methyl groups. Subsequent SELEX experiments refined the model, stipulating that high-affinity interaction with methylated DNA was facilitated if the methylated CpG dinucleotide was flanked by A/T base pairs on each side (see Table 13.1) (Klose et al. 2005; Ghosh et al. 2010). Exploration of the MeCP2 interaction with chromatin demonstrated that the protein has a very high affinity for nucleosomes, equivalent to that of linker histone. It is able to mediate chromatin compaction in vitro, a function requiring residues outside the MBD domain (Georgel et al. 2003).

MeCP2 shows relatively high expression levels in the brain, particularly in neurons, where its levels approach that of histone octamers (Skene et al. 2010). The influence of MeCP2 on normal chromatin architecture in neurons is underscored by the phenotypes of individuals deficient in MeCP2 function. Mutations in the human MeCP2 gene result in the neurodevelopmental disorder Rett syndrome (Amir et al. 1999; Chen et al. 2001; Guy et al. 2001). However, despite the important role(s) in development and disease evident in individuals harboring mutations, MeCP2

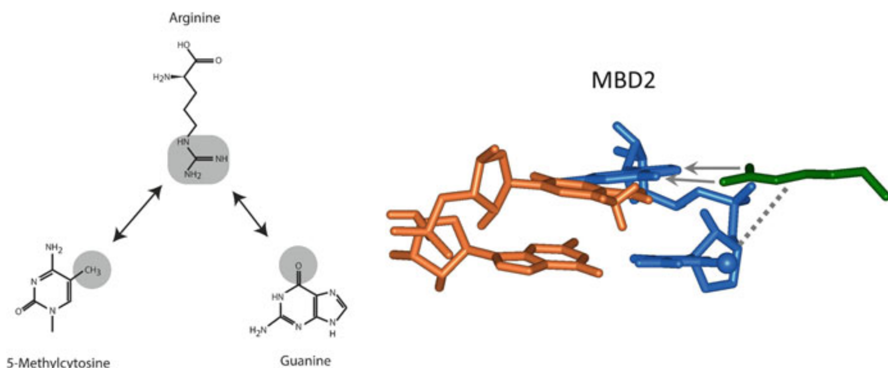


Fig. 2 Schematic representation of the 5mC-Arg-Gua triad. The guanidinium group of arginine forms a triad with the methylated C and its adjacent G residue. The figure depicts a van der Waals contact with the cytosine methyl group and hydrogen bonds with guanine. In the right side, this interaction is illustrated in the MBD complex (2KY8). Arg24 is shown in *green*, the CpG site in *blue* and *orange*

Table 1 Known DNA sequence preferences of MBPs

MeCP2	MG-(A/T) $_{\geq 4}$ [16]
MBD1	TMGCA [25]
	TGMGCA [25]
MBD2	CMGG [15]
Kaiso	MGMG [87]
	TCCTGCNA [88]
ZFP57	TGCMGC [93]
KLF4	(A/G)(G/A)GG(M/T)G(C/T) [99]
	GGG(M/T)G(T/G)GG [100]
EGR1, WT1	GMG(T/G)GGGMG [104]

M stands for 5-methylcytosine, N stands for any nucleotide

deletion in mice has minimal impact on global gene regulation. Thus, detailed mechanistic insights into how disruption of MeCP2 causes developmental failure or Rett syndrome are currently lacking. Recently, it has been suggested that MeCP2 recognizes methylated CpA dinucleotides within the gene body of long genes; surprisingly, those long genes are downregulated when MeCP2 is absent (Gabel et al. 2015). Although it is still unclear if this observation has relevance to tissues other than the brain (methylation in non-CpG contexts is typically found at high levels in embryonic stem cells or the brain), this emerging data suggest a novel paradigm explaining how MeCP2 regulates gene expression.

2.2 MBD1

MBD1, like other MBD proteins, was initially discovered in homology screens using the MeCP2 MBD as template. MBD1 is distinctive among the MBD proteins in that, in addition to the MBD, the protein has either two or three CXXC zinc finger domains; the isoforms resulting from alternative splicing have different numbers of zinc finger domains (Fujita et al. 1999). Like MeCP2, MBD1 also contains a transcriptional repression domain (TRD) near its carboxyl-terminus (Fig. 1). The structure of the methyl-CpG-binding domain of MBD1 in complex with methylated DNA was solved by Shirakawa and colleagues in 2001 (Ohki et al. 2001). This seminal work demonstrated that while the DNA substrate is completely symmetrical, the protein interface recognizing this sequence is not. A pair of beta strands lies in the major groove of DNA with specific chemical contacts formed by amino acid side chains (Fig. 3). One cytosine methyl group sits in a pocket formed by specific valine, tyrosine, and arginine residues. The other methyl group is recognized by a second arginine residue (reminiscent of the 5mC-Arg-Gua triad in MeCP2) and a serine (Ohki et al. 2001). Interestingly, DNA binding by the MBD1 MBD (and presumably other MBD family members) causes an unstructured loop to fold into a highly structured interaction with the DNA backbone on one strand. Biochemical analysis indicated that, like in the case of MeCP2, nucleotides flanking the

Fig. 3 Schematic representation of the methyl-CpG recognition by MBD1. The symmetrical DNA substrate – methylated CpG paired with methylated CpG – is recognized by an asymmetric protein recognition module. Two separate units interact with the methylated CpG moiety on each strand, V20-R22-Y34 for the binding surface for one strand and R44-S45 for the other strand. Both arginine residues, R22 and R44, form 5mC-Arg-Gua triads. On the right side, the clamping interaction of both arginine residues (shown in *light green*) into the DNA helix in MBD1 (1IG4) is illustrated. The protein is shown as ribbon tube in *dark green* and the DNA in *blue* and *orange*

methylated CpG dinucleotide exert an influence on binding affinity of the MBD1 MBD domain. An adenine residue at position +2 relative to the methylated CpG and a thymine at position −2 provided a three- to fivefold increase in binding affinity (Table 13.1) (Clouaire et al. 2010). Interestingly, the flanking residues providing increased binding affinity to the MBD1 MBD domain actually decreased binding interactions with the MBD domain from MeCP2, suggesting that nucleotides flanking a methylated CpG have a strong influence on which MBD protein may stably interact with a given sequence.

In addition to the MBD, the zinc fingers of MBD1 add an additional DNA-binding interface. While MBD1 isoforms with all three zinc fingers can repress genes regardless of their promoter methylation status, MBD1 lacking the third zinc finger, CxxC3, can only suppress gene expression when the promoter is methylated, suggesting the CxxC3 is essential for binding to unmethylated templates (Fujita et al. 2000; Jørgensen et al. 2004). The extent to which the zinc fingers impact on the target specificity of MBD1 was unclear until recently. In embryonic stem cells, biotin tagged MBD1 localized at highly methylated regions, and the enrichment was lost upon depletion of DNA methylation. In addition, the targeting of MBD1 to unmethylated templates was observed only when the MBD was deleted, suggesting the recruiting mechanism of MBD1 is dominated by the MBD-methyl-CpG interaction (Baubec et al. 2013).

2.3 MBD2

MBD2 is a member of a chromatin remodeling complex, Mi-2/NuRD (Nucleosome Remodeling Deacetylase), which is proposed to act as a repressor connecting DNA methylation and histone deacetylation (Feng and Zhang 2001). Like MeCP2 and

MBD1, MBD2 contains a TRD in addition to an MBD. MBD2 also has a glycine-arginine (GR) repeat and a coiled-coil (CC) domain (Fig. 1), which is essential for binding to the Mi-2/NuRD complex (Gnanapragasam et al. 2011). MBD2 was also reported to be an integral component of the MeCP1 complex which was subsequently shown to have biochemical similarities to Mi-2/NuRD (Ng et al. 1999; Zhang et al. 1999).

The interaction of the MBD2 methyl-CpG-binding domain with its DNA substrate has been defined in detail for chicken MBD2 (which is >95 % identical to human) by Williams and colleagues (Scarsdale et al. 2011). The DNA substrate initially employed for this set of analyses was derived from a chicken globin gene and provided advantages not evident in model DNA substrates used in previous structural studies of MBD proteins. It was observed that MBD2 bound the DNA substrate in a single orientation that was largely determined by a guanine base at position +1 relative to the methylated CpG (Table 13.1). Further, mutation of this residue to other bases reduced the binding affinity by up to tenfold. Similarly as for MeCP2 and MBD1, base-specific contacts with the methylated CpG palindrome are mediated by a pair of arginine residues (R24 and R46), as well as a tyrosine residue (Y36). The molecular basis for the preference of MBD2 for mCGG sequences likely lies in base-specific contacts formed between a lysine residue (K32) and the +1 guanine base.

MBD2 is expressed in most somatic cells and is particularly abundant in embryonic stem cells in mice (Hendrich and Bird 1998). Despite this intriguing expression pattern, MBD2 knockout mice are viable and fertile (Hendrich et al. 2001). MBD2 has two predominant isoforms, MBD2a and MBD2c; MBD2c lacks the carboxyl-terminal region including the coiled-coil domain integral to the interaction with Mi-2/NuRD complex (Hendrich and Bird 1998). Recently, differential expression of MBD2a and MBD2c was shown in human pluripotent stem cells (hPSCs); MBD2c is dominant in hPSCs, while MBD2a is dominant in fibroblasts (Lu et al. 2014). Interestingly, MBD2a, but not MBD2c, can interact with NuRD complex and promote differentiation, while MBD2c enhances reprogramming efficiency when overexpressed in fibroblasts.

2.4 MBD3

As predicted from its high sequence similarity with MBD2 (Fig. 1), MBD3 is also a member of Mi-2/NuRD, although MBD2 and MBD3 form mutually exclusive complexes (Le Guezennec et al. 2006). In contrast to the mild phenotypes of the MBD2 knockout mice, MBD3 deletion causes early embryonic lethality. Furthermore, MBD3-null ES cells can maintain stemness, even in the absence of leukemia inhibitory factor (LIF) (Hendrich et al. 2001; Kaji et al. 2006). These striking phenotypes suggest that MBD2 and MBD3 have nonredundant role(s). Importantly, deletion of MBD3 can significantly enhance reprogramming efficiency, suggesting that MBD3 functions as a barrier for reprogramming (Luo et al. 2013; Rais et al. 2013). However, the field is not in a complete agreement with the role of MBD3 in reprogramming, suggesting a possibility that the role of MBD3 is highly context dependent (dos Santos et al. 2014; Bertone et al. 2015; Zviran et al. 2015). Moreover, in uterine serous carcinoma patients, a small segment of

chromosome 19 containing MBD3 is frequently deleted, suggesting a critical role of MBD3 in tumorigenesis or tumor progression (Zhao et al. 2013).

MBD3 from mammals lacks the capacity for high-affinity interaction with methylated DNA in conventional biochemical assays (Fraga et al. 2003; Hendrich and Bird 1998). In contrast, the amphibian protein displays a strong preference for methylated substrates (Wade et al. 1999). Close inspection of the amino acid sequence of MBD3 MBD domains from multiple species reveals that mammals differ from Amphibia, fish, reptiles, and birds at two critical positions encoding the conserved tyrosine residue involved in specific contacts with a 5-methyl C methyl group (changed to phenylalanine in mammals) and a lysine residue (changed to histidine in mammals) that contacts the +1 guanine base in MBD2. These amino acid changes are consistent with the binding behavior observed in classic biochemical assays.

High resolution NMR has the ability to observe interactions that may be missed with other techniques. In the case of MBD3, a precise structural determination was not pursued, but the interaction of human MBD3 with various modified DNA substrates was probed (Cramer et al. 2014). MBD3 has a modest preference for methylated DNA in NMR experiments; its preference is not as marked as MBD2. The protein also appears to lack the capacity to recognize oxidized forms of 5-methyl cytosine.

Despite the accumulation of evidence on the critical function of MBD3 in biology, the detailed molecular mechanism of how MBD3, or MBD3/NuRD, regulates gene expression and chromatin structure is still unclear. Unlike MBD2, MBD3 does not show strong binding preference to methylated DNA (Saito and Ishikawa 2002; Fraga et al. 2003). Indeed, most cells express a splice variant of MBD3 that disrupts the canonical MBD sequence (Hendrich and Bird 1998; Hendrich and Tweedie 2003). Genomic localization analyses revealed that MBD3 preferentially localizes at unmethylated CpG-rich regions, including CpG islands, while MBD2 distributes across the genome in a methylation-dependent manner (Baubec et al. 2013). Surprisingly, MBD3 was reported to preferentially bind to 5hmC and localize in a TET1 (ten-eleven translocation 1)-dependent manner, consistent with a direct interaction with oxidized forms of 5-methyl C (Yildirim et al. 2011). While this report was initially attractive in concept, it has not been supported by subsequent biochemical and genomic experiments (Hashimoto et al. 2012a; Baubec et al. 2013; Cramer et al. 2014).

2.5 MBD4

MBD4 contains two functional domains, an MBD and a glycosylase domain (GD), separated by a relatively long region of unknown function (Fig. 1). This presence of a glycosylase domain positions MBD4 as a unique member among the MBD family. The MBD domain has high affinity for methylated CpG-containing substrates; it has similar affinity for the deamination product of that substrate – methylated CpG base paired (mismatch) with TpG (Hendrich et al. 1999; Otani et al. 2013). Not surprisingly, MBD4 possesses enzymatic activity that can repair mCpG/TpC or mCpG/hmUpG double-strand mismatches (Hendrich and Bird 1998; Hendrich et al. 1999). MBD4 knockout mice are viable and fertile with minor phenotypes,

including a slight increase in C to T mutations at CpG sites (Millar et al. 2002; Wong et al. 2002). Although the deletion of MBD4 itself does not impact tumorigenesis, it increases tumor incidence in a susceptible genetic background (mutation in the adenomatous polyposis coli (APC) gene). In addition, mutations in MBD4 have been observed in human colorectal tumors with microsatellite instability (Riccio et al. 1999). Taken together, these observations suggest that MBD4 may play an important role in tumor progression by regulating DNA mismatch repair.

The crystal structures of the MBD of MBD4 with a methylated DNA substrate revealed that MBD4 recognizes these templates using the 5mC-Arg-Gua triad in a manner highly similar to MeCP2 and other MBD family members (Otani et al. 2013). While there are some minor alterations, the overall structure of the MBD4 MBD domain bound to a mismatched substrate is highly similar to that of the protein bound to symmetric methylated CpG-containing DNA. In contrast to the MBD, the glycosylase domain of MBD4 binds to DNA containing a G:X mismatch in a very different manner (Hashimoto et al. 2012b). The target nucleotide is flipped out from the DNA strand and an arginine fills that space. The flipped base is associated with the active-site cleft. Importantly, the crystal structure of full-length MBD4 containing both the MBD and GD is yet to be solved, and it is unclear to what level the two domains communicate. This question has been partially approached via analysis of solution structure; although MBD4 shows a slow exchange rate between different DNA molecules (intermolecular exchange), it has rapid exchange rate between the two binding sites on the same dsDNA molecule (intramolecular exchange) (Walavalkar et al. 2014). These data support a local hopping model in which the MBD of MBD4 rapidly scans multiple methylated CpG sites and supports the mismatch repair conducted by the GD.

2.6 MBD5 and MBD6

MBD5 and MBD6 are relatively uncharacterized members of the MBD family; biochemical and structural analyses on their methylcytosine-binding ability are limited. MBD5 and MBD6 are known to be associated with some neurodevelopmental disorders (Cukier et al. 2012; Talkowski et al. 2012). MBD5-null mice show growth defects and exhibit preweaning lethality, showing several characteristic phenotypes observed in 2q23.1 microdeletion syndrome patients (Du et al. 2012). Like other MBD proteins, MBD5 and MBD6 localize at heterochromatic regions in an MBD-dependent manner (Laget et al. 2010). However, EMSA experiments using purified MBDs from MBD5 and MBD6 indicate that these domains display no methylcytosine-binding capacity. This property may be explained by the loss of a characteristic loop structure, which is critical for DNA backbone contacts in the MBDs of MBD5 and MBD6 (Hendrich and Tweedie 2003). Interestingly, although the incomplete MBDs of these proteins have lost methylcytosine-binding affinity, they can interact with mammalian PR-DUB polycomb protein complex, which is known as a histone H2A deubiquitinase (Baymaz et al. 2014). These distinct differences from other MBD proteins may assign a specialized function of MBD5 and MBD6.

3 SET- and RING-Associated (SRA) Domain

3.1 UHRF1 (Ubiquitin-Like with PHD and Ring Finger Domains)

UHRF1 (ubiquitin-like, containing PHD and RING finger domains 1, also known as ICBP90 or Np95) contains ubiquitin-like (UBL), tandem tudor, plant homeodomain (PHD), SET- and RING-associated (SRA), and really interesting new gene (RING) domains (Fig. 1) and was originally identified as a potential regulator for topoisomerase IIα (Hopfner et al. 2000). Overexpression of UHRF1 is observed in various types of tumors and drives tumorigenesis by inducing genome-wide DNA hypomethylation (Mudbhary et al. 2014). UHRF1 is essential for maintenance of proper DNA methylation levels by recruiting DNA methyltransferase 1 (DNMT1) to replication foci (Bostick et al. 2007; Sharif et al. 2007). Deletion of UHRF1 in mice causes genome-wide DNA hypomethylation and results in embryonic lethality, presumably due to the dysfunction of DNMT1.

UHRF1 recognizes hemimethylated DNA with its SRA domain using an interesting mechanism, referred to as base flipping, commonly found in DNA methyltransferases (Arita et al. 2008; Hashimoto et al. 2008; Avvakumov et al. 2008). To achieve this base-flipping recognition mechanism, the SRA domain encodes two different loop structures responsible for CpG recognition and base flipping. The two loops capture DNA from opposite directions; the CpG recognition loop and the base-flipping loop capture the major groove and minor groove of DNA, respectively. The flipped-out methylcytosine is stabilized in a binding pocket with planar stacking contacts, Watson-Crick polar hydrogen bonds, and van der Waals interactions. The use of this base-flipping mechanism uniquely positions UHRF proteins in the MBP family; the SRA domain is the first domain which conducts base flipping without enzymatic activity (Song et al. 2012).

In addition to hemimethylated DNA, UHRF1 also recognizes histone modifications, such as H3K9me3, unmodified H3K9, and H3R2 through its histone reader domains (Rajakumara et al. 2011; Hu et al. 2011; Nady et al. 2011). This interaction is allosterically regulated by phosphatidylinositol 5-phospate (PI5P), which alters the local structure around the tudor and PHD domains (Gelato et al. 2014). Interestingly, it has been demonstrated that the interaction between UHRF1 and H3K9me3 is essential for the maintenance of DNA methylation, suggesting a role for UHRF1 as a molecular hub connecting DNA methylation and histone modifications (Rothbart et al. 2012; Liu et al. 2013a).

3.2 UHRF2

UHRF2 (also known as NIRF) has high sequence similarity with its paralog UHRF1 (Fig. 1). They are the only two proteins with an SRA domain in humans (Mori et al. 2002). Based on their sequence similarity, UHRF2 and UHRF1 likely share functions; UHRF2 also recognizes hemimethylated DNA and interacts with DNMT1 (Zhang et al. 2011). However, there are critical differences between UHRF1 and

UHRF2. Most importantly, unlike UHRF1, UHRF2, through the SRA domain, can specifically bind to 5hmC using base flipping (Zhou et al. 2014). In addition, UHRF2 and UHRF1 are differentially expressed; UHRF2 is downregulated in embryonic stem cells and gradually upregulated upon differentiation, whereas UHRF1 shows an opposite pattern (Pichler et al. 2011). Moreover, the introduction of UHRF2 into UHRF1-null embryonic stem cells cannot rescue the hypomethylation phenotype, suggesting a differential functionality of UHRF2, at least in embryonic stem cells (Zhang et al. 2011). With these similarities and differences, it is yet unclear how UHRF1 and UHRF2 cooperatively (or distinctively) function in cells; analysis with UHRF2 knockout mice would provide more information.

4 Transcription Factors

4.1 Kaiso

Being a member of the BTB/POZ (*B*road complex, *T*ramtrack, and *B*ric-a-brac/*po*xvirus and *z*inc finger) family, Kaiso contains three C2H2 zinc fingers and was originally identified as a binding partner of p120 catenin (Daniel and Reynolds 1999). In addition to p120 catenin, Kaiso also interacts with a repression complex, N-CoR (nuclear receptor corepressor), and suppresses the expression of MTA2, a member of Mi-2/NuRD complex, in a methylation-dependent manner (Yoon et al. 2003). As N-CoR complex contains histone deacetylase, the recruitment of N-CoR complex mediated by Kaiso is proposed as a potential mechanism of DNA methylation-dependent gene repression. However, Kaiso is also known to associate with p53 and upregulate apoptosis-related genes, suggesting pleiotropic roles of Kaiso in different biological contexts (Koh et al. 2014, 2015). Moreover, Kasio showed both pro- and antitumorigenic activities, which also position Kaiso as a context-dependent regulator (Prokhortchouk et al. 2006; Soubry et al. 2010; Koh et al. 2014).

Kaiso has preferential sequence determinants for binding: two consecutive methylated CpG dinucleotides or to a chemically similar, albeit unmethylated counterpart, TCCTGCCA (Prokhortchouk et al. 2001; Daniel et al. 2002). The crystal structures of Kaiso have been solved using two templates: a methylated template, MeECad (promoter region of E-cadherin) containing two methylated CpG dinucleotides, and an unmethylated sequence TCCTGCCA (Buck-Koehntop et al. 2012). The two structures were almost identical; Kaiso recognizes the methyl group, either in mCpG or TpG dinucleotides, using the 5mC-Arg-Gua triad structure. The first two zinc fingers hold the major groove of DNA, and the third zinc finger (together with the C-terminal extension) enables high-affinity binding.

4.2 Zfp57

Zfp57 belongs to the KRAB (Kruppel-associated box) zinc finger family. Zfp57, together with its binding cofactor KAP1, regulates imprinted genes by targeting CpG-rich regions, known as imprinting control regions (ICRs) (Li et al. 2008).

Loss of Zfp57 in mice results in early embryonic lethality, presumably due to loss of the methylation imprint (Li et al. 2008). Interestingly, the loss of DNA methylation imprints in Zfp57-null embryonic stem cells cannot be rescued by supplementation with exogenous Zfp57, suggesting a maintenance role in genomic imprinting (Zuo et al. 2012). In addition, loss-of-function mutations in human Zfp57 are associated with global imprinting disorder, transient neonatal diabetes (TND), suggesting a critical role of Zfp57 both in mice and humans (Mackay et al. 2008). Zfp57 recognizes the TGCCGC motif containing a methylated CpG dinucleotide with strong preference for fully methylated template (Quenneville et al. 2011; Liu et al. 2012). Like MBD proteins and Kaiso, Zfp57 uses the 5mC-Arg-Gua triad for methylated cytosine recognition.

4.3 KLF4

KLF4 is a member of the Kruppel-like factor family and contains three standard Kruppel-like zinc fingers. KLF4 is probably best known as a Yamanaka factor for reprogramming somatic cells to induced pluripotency (Takahashi and Yamanaka 2006). In addition to reprogramming, KLF4 is required for normal skin and colon development, and KLF4 knockout mice die soon after birth (Segre et al. 1999; Katz et al. 2002). KLF4 recognizes both methylated and unmethylated DNA templates using the 5mC-Arg-Gua triad with a modest difference in binding affinity (Liu et al. 2014; Shields and Yang 1998; Chen et al. 2008). KLF4 progressively loses binding affinity as 5mC is sequentially oxidized into 5hmC, 5fC, and 5caC. Interestingly, genome-wide studies showed that about half of KLF4-binding sites in vivo are highly methylated (Hu et al. 2013).

4.4 EGR1 and WT1

EGR1 (early growth response protein 1) and WT1 (Wilms' tumor 1) are two related C2H2 type zinc finger proteins, both recognizing the same 9-bp consensus sequence, GCG(T/G)GGGCG (Pavletich and Pabo 1991; Stoll et al. 2007). Consistent with other MBPs, these two proteins interact with a methylated DNA template using the 5mC-Arg-Gua triad (Hashimoto et al. 2014). Like KLF4, EGR1 and WT1 have similar binding affinity to their consensus DNA sequence regardless of its methylation status. Interestingly, Hashimoto et al. clarified that EGR1, but not WT1, loses binding capacity when the cytosine within the motif is oxidized into 5caC, suggesting the potential for epigenetic regulation triggered by the oxidation of methylated cytosine.

Conclusion

In this chapter, we reviewed the research on proteins that recognize methylated DNA. Several themes have emerged from recent biochemical, structural, and genomic studies of MBPs. With a single exception (the SRA domain-containing

proteins), all MBP proteins characterized to date utilize a version of the 5-methyl C, arginine, and guanine triad in molecular recognition of methylated CpG. In many MBPs, a pair of arginine residues makes contact with the methylated CpGs on the two DNA strands. This near-universal mechanism is somewhat surprising given that outside of this feature the various families have little sequence or structural similarity.

The recognition of the symmetrically methylated CpG double-stranded DNA moiety is universally performed by MBP proteins that lack symmetry in their DNA-binding domains. Thus, the amino acids recognizing the methylated CpG in one strand differ from those recognizing the other strand. This feature implies that some directionality of binding occurs at methylated loci in vivo. The downstream outcomes of this unanticipated behavior are currently unknown, but it is tempting to speculate that orientation relative to genomic features may have critical impact(s) on orientation of functional domains and interaction partners that influence local chromatin features.

Detailed analyses of structural and biochemical data have also indicated that the MBD family members have the capacity to recognize DNA features outside the methylated CpG sequence. While genomic mapping experiments to localize these proteins in living cells are in their infancy, it seems likely that some level of functional specificity may be produced by the preference of individual MBD family members for chemical information found flanking the methylated CpG. To what extent cells utilize this differential capacity to distinguish bases flanking a methylated CpG sequence to specify recruitment of a unique MBP (and associated cofactors) is incompletely understood.

As a dinucleotide sequence (i.e., meCpG) lacks the chemical information inherent in the longer binding sites typical of most transcription factors, it has been puzzling how the various MBP proteins would be distributed across the genome. The structural and biochemical work reviewed here places this problem in a new context. Most MBP proteins utilize a common protein feature (arginine) to recognize chemical features of the methylated CpG dinucleotide. Additional protein functional groups recognize flanking sequence information, suggesting that the true consensus recognition sequence for these proteins is not a simple methylated CpG, but the methylated CpG with its flanking sequence context that differs for each MBP (Table 13.1). This feature – recognition of a common modified dinucleotide flanked by specific sequence information – endows the MBP proteins with considerable flexibility to respond to DNA methylation with different outputs at unique loci within the genome. Unraveling how cells utilize this surprising flexibility to resolve epigenetic regulation remains a principal challenge of current genetic and genomic experiments.

Acknowledgments The authors gratefully acknowledge the members of the Wade laboratory for many useful discussions throughout the completion of this work and A. Jeltsch for providing the structural views shown in Figs. 2 and 3. This work was supported by the Intramural Research Program of the National Institute of Environmental Health Sciences, NIH (ES101965 to P.A.W.).

References

Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat Genet. 1999;23(2):185–8.

Arita K, Ariyoshi M, Tochio H, Nakamura Y, Shirakawa M. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature. 2008;455(7214):818–21.

Avvakumov GV, Walker JR, Xue S, Li Y, Duan S, Bronner C, et al. Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. Nature. 2008;455(7214):822–5.

Baubec T, Ivánek R, Lienert F, Schübeler D. Methylation-dependent and-independent genomic targeting principles of the MBD protein family. Cell. 2013;153(2):480–92.

Baymaz H, Fournier A, Laget S, Ji Z, Jansen PW, Smits AH, et al. MBD5 and MBD6 interact with the human PR-DUB complex through their methyl-CpG-binding domain. Proteomics. 2014;14(19):2179–89.

Bertone P, Hendrich B, Silva JC. Mbd3 and deterministic reprogramming. bioRxiv. 2015:013904.

Bird A. DNA methylation patterns and epigenetic memory. Genes Dev. 2002;16(1):6–21.

Bostick M, Kim JK, Estève P-O, Clark A, Pradhan S, Jacobsen SE. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science. 2007;317(5845):1760–4.

Buck-Koehntop BA, Stanfield RL, Ekiert DC, Martinez-Yamout MA, Dyson HJ, Wilson IA, et al. Molecular basis for recognition of methylated and specific DNA sequences by the zinc finger protein Kaiso. Proc Natl Acad Sci. 2012;109(38):15229–34.

Chen RZ, Akbarian S, Tudor M, Jaenisch R. Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. Nat Genet. 2001;27(3):327–31.

Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell. 2008;133(6):1106–17.

Clouaire T, de las Heras JI, Merusi C, Stancheva I. Recruitment of MBD1 to target genes requires sequence-specific interaction of the MBD domain with methylated DNA. Nucleic Acids Res. 2010;38(14):4620–34.

Cramer JM, Scarsdale JN, Walavalkar NM, Buchwald WA, Ginder GD, Williams DC. Probing the dynamic distribution of bound states for methylcytosine-binding domains on DNA. J Biol Chem. 2014;289(3):1294–302.

Cukier HN, Lee JM, Ma D, Young JI, Mayo V, Butler BL, et al. The expanding role of MBD genes in autism: identification of a MECP2 duplication and novel alterations in MBD5, MBD6, and SETDB1. Autism Res. 2012;5(6):385–97.

Daniel JM, Reynolds AB. The catenin p120(ctn) interacts with Kaiso, a novel BTB/POZ domain zinc finger transcription factor. Mol Cell Biol. 1999;19(5):3614–23.

Daniel JM, Spring CM, Crawford HC, Reynolds AB, Baig A. The p120ctn-binding partner Kaiso is a bi-modal DNA-binding protein that recognizes both a sequence-specific consensus and methylated CpG dinucleotides. Nucleic Acids Res. 2002;30(13):2911–9.

Domcke S, Bardet AF, Ginno PA, Hartl D, Burger L, Schübeler D. Competition between DNA methylation and transcription factors determines binding of NRF1. Nature. 2015;528:575–79.

dos Santos RL, Tosti L, Radzisheuskaya A, Caballero IM, Kaji K, Hendrich B, et al. MBD3/NuRD facilitates induction of pluripotency in a context-dependent manner. Cell Stem Cell. 2014; 15(1):102–10.

Du Y, Liu B, Guo F, Xu G, Ding Y, Liu Y, et al. The essential role of Mbd5 in the regulation of somatic growth and glucose homeostasis in mice. PLoS One. 2012;7(10):e47358.

Feng Q, Zhang Y. The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes. Genes Dev. 2001;15(7):827–32.

Fraga MF, Ballestar E, Montoya G, Taysavang P, Wade PA, Esteller M. The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. Nucleic Acids Res. 2003;31(6):1765–74.

Fujita N, Takebayashi S-i, Okumura K, Kudo S, Chiba T, Saya H, et al. Methylation-mediated transcriptional silencing in euchromatin by methyl-CpG binding protein MBD1 isoforms. Mol Cell Biol. 1999;19(9):6415–26.

Fujita N, Shimotake N, Ohki I, Chiba T, Saya H, Shirakawa M, et al. Mechanism of transcriptional regulation by methyl-CpG binding protein MBD1. Mol Cell Biol. 2000;20(14):5107–18.
Gabel HW, Kinde B, Stroud H, Gilbert CS, Harmin DA, Kastan NR, et al. Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. Nature. 2015;522(7554):89–93.
Gelato KA, Tauber M, Ong MS, Winter S, Hiragami-Hamada K, Sindlinger J, et al. Accessibility of different histone H3-binding domains of UHRF1 is allosterically regulated by phosphatidylinositol 5-phosphate. Mol Cell. 2014;54(6):905–19.
Georgel PT, Horowitz-Scherer RA, Adkins N, Woodcock CL, Wade PA, Hansen JC. Chromatin compaction by human MeCP2 assembly of novel secondary chromatin structures in the absence of DNA methylation. J Biol Chem. 2003;278(34):32181–8.
Ghosh RP, Horowitz-Scherer RA, Nikitina T, Shlyakhtenko LS, Woodcock CL. MeCP2 binds cooperatively to its substrate and competes with histone H1 for chromatin binding sites. Mol Cell Biol. 2010;30(19):4656–70.
Gnanapragasam MN, Scarsdale JN, Amaya ML, Webb HD, Desai MA, Walavalkar NM, et al. p66α–MBD2 coiled-coil interaction and recruitment of Mi-2 are critical for globin gene silencing by the MBD2–NuRD complex. Proc Natl Acad Sci. 2011;108(18):7487–92.
Guy J, Hendrich B, Holmes M, Martin JE, Bird A. A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. Nat Genet. 2001;27(3):322–6.
Hashimoto H, Horton JR, Zhang X, Bostick M, Jacobsen SE, Cheng X. The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. Nature. 2008;455(7214):826–9.
Hashimoto H, Liu Y, Upadhyay AK, Chang Y, Howerton SB, Vertino PM, et al. Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Res. 2012a;40(11):4841–9.
Hashimoto H, Zhang X, Cheng X. Excision of thymine and 5-hydroxymethyluracil by the MBD4 DNA glycosylase domain: structural basis and implications for active DNA demethylation. Nucleic Acids Res. 2012b;40(17):8276–84.
Hashimoto H, Olanrewaju YO, Zheng Y, Wilson GG, Zhang X, Cheng X. Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence. Genes Dev. 2014;28(20):2304–13.
Hendrich B, Bird A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol Cell Biol. 1998;18(11):6538–47.
Hendrich B, Tweedie S. The methyl-CpG binding domain and the evolving role of DNA methylation in animals. Trends Genet. 2003;19(5):269–77.
Hendrich B, Hardeland U, Ng H-H, Jiricny J, Bird A. The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. Nature. 1999;401(6750):301–4.
Hendrich B, Guy J, Ramsahoye B, Wilson VA, Bird A. Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. Genes Dev. 2001;15(6):710–23.
Ho KL, McNae IW, Schmiedeberg L, Klose RJ, Bird AP, Walkinshaw MD. MeCP2 binding to DNA depends upon hydration at methyl-CpG. Mol Cell. 2008;29(4):525–31.
Hopfner R, Mousli M, Jeltsch J-M, Voulgaris A, Lutz Y, Marin C, et al. ICBP90, a novel human CCAAT binding protein, involved in the regulation of topoisomerase IIα expression. Cancer Res. 2000;60(1):121–8.
Hu L, Li Z, Wang P, Lin Y, Xu Y. Crystal structure of PHD domain of UHRF1 and insights into recognition of unmodified histone H3 arginine residue 2. Cell Res. 2011;21(9):1374–8.
Hu S, Wan J, Su Y, Song Q, Zeng Y, Nguyen HN, et al. DNA methylation presents distinct binding sites for human transcription factors. Elife. 2013;2:e00726.
Jørgensen HF, Ben-Porath I, Bird AP. Mbd1 is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains. Mol Cell Biol. 2004;24(8):3387–95.
Kaji K, Caballero IM, MacLeod R, Nichols J, Wilson VA, Hendrich B. The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. Nat Cell Biol. 2006;8(3):285–92.
Katz JP, Perreault N, Goldstein BG, Lee CS, Labosky PA, Yang VW, et al. The zinc-finger transcription factor Klf4 is required for terminal differentiation of goblet cells in the colon. Development. 2002;129(11):2619–28.
Klose RJ, Sarraf SA, Schmiedeberg L, McDermott SM, Stancheva I, Bird AP. DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. Mol Cell. 2005;19(5):667–78.

Koh D-I, Han D, Ryu H, Choi W-I, Jeon B-N, Kim M-K, et al. KAISO, a critical regulator of p53-mediated transcription of CDKN1A and apoptotic genes. Proc Natl Acad Sci. 2014;111(42):15078–83.

Koh D-I, An H, Kim M-Y, Jeon B-N, Choi S-H, Hur SS, et al. Transcriptional activation of APAF1 by KAISO (ZBTB33) and p53 is attenuated by RelA/p65. Biochim Biophys Acta. 2015;1849(9):1170–8.

Kohli RM, Zhang Y. TET enzymes, TDG and the dynamics of DNA demethylation. Nature. 2013;502(7472):472–9.

Kriaucionis S, Tahiliani M. Expanding the epigenetic landscape: novel modifications of cytosine in genomic DNA. Cold Spring Harb Perspect Biol. 2014;6(10):a018630.

Laget S, Joulie M, Le Masson F, Sasai N, Christians E, Pradhan S, et al. The human proteins MBD5 and MBD6 associate with heterochromatin but they do not bind methylated DNA. PLoS One. 2010;5(8):e11982.

Le Guezennec X, Vermeulen M, Brinkman AB, Hoeijmakers WA, Cohen A, Lasonder E, et al. MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. Mol Cell Biol. 2006;26(3):843–51.

Lewis JD, Meehan RR, Henzel WJ, Maurer-Fogy I, Jeppesen P, Klein F, et al. Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell. 1992;69(6):905–14.

Li X, Ito M, Zhou F, Youngson N, Zuo X, Leder P, et al. A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. Dev Cell. 2008;15(4):547–57.

Liu Y, Toh H, Sasaki H, Zhang X, Cheng X. An atomic model of Zfp57 recognition of CpG methylation within a specific DNA sequence. Genes Dev. 2012;26(21):2374–9.

Liu X, Gao Q, Li P, Zhao Q, Zhang J, Li J, et al. UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. Nat Commun. 2013a;4:1563.

Liu Y, Zhang X, Blumenthal RM, Cheng X. A common mode of recognition for methylated CpG. Trends Biochem Sci. 2013b;38(4):177–83.

Liu Y, Olanrewaju YO, Zheng Y, Hashimoto H, Blumenthal RM, Zhang X, et al. Structural basis for Klf4 recognition of methylated DNA. Nucleic Acids Res. 2014;42(8):4859–67.

Lu Y, Loh Y-H, Li H, Cesana M, Ficarro SB, Parikh JR, et al. Alternative splicing of MBD2 supports self-renewal in human pluripotent stem cells. Cell Stem Cell. 2014;15(1):92–101.

Luo M, Ling T, Xie W, Sun H, Zhou Y, Zhu Q, et al. NuRD blocks reprogramming of mouse somatic cells into pluripotent stem cells. Stem Cells. 2013;31(7):1278–86.

Mackay DJ, Callaway JL, Marks SM, White HE, Acerini CL, Boonen SE, et al. Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. Nat Genet. 2008;40(8):949–51.

Meehan RR, Lewis JD, McKay S, Kleiner EL, Bird AP. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell. 1989;58(3):499–507.

Meehan R, Lewis JD, Bird AP. Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. Nucleic Acids Res. 1992;20(19):5085–92.

Millar CB, Guy J, Sansom OJ, Selfridge J, MacDougall E, Hendrich B, et al. Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. Science. 2002;297(5580):403–5.

Mori T, Li Y, Hata H, Ono K, Kochi H. NIRF, a novel RING finger protein, is involved in cell-cycle regulation. Biochem Biophys Res Commun. 2002;296(3):530–6.

Mudbhary R, Hoshida Y, Chernyavskaya Y, Jacob V, Villanueva A, Fiel MI, et al. UHRF1 overexpression drives DNA hypomethylation and hepatocellular carcinoma. Cancer Cell. 2014;25(2):196–209.

Nady N, Lemak A, Walker JR, Avvakumov GV, Kareta MS, Achour M, et al. Recognition of multivalent histone states associated with heterochromatin by UHRF1 protein. J Biol Chem. 2011;286(27):24300–11.

Nan X, Meehan RR, Bird A. Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. Nucleic Acids Res. 1993;21(21):4886–92.

Ng H-H, Zhang Y, Hendrich B, Johnson CA, Turner BM, Erdjument-Bromage H, et al. MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. Nat Genet. 1999;23(1):58–61.

Ohki I, Shimotake N, Fujita N, Jee J-G, Ikegami T, Nakao M, et al. Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. Cell. 2001;105(4):487–97.

Otani J, Arita K, Kato T, Kinoshita M, Kimura H, Suetake I, et al. Structural basis of the versatile DNA recognition ability of the methyl-CpG binding domain of methyl-CpG binding domain protein 4. J Biol Chem. 2013;288(9):6351–62.

Pavletich NP, Pabo CO. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science. 1991;252(5007):809–17.

Pichler G, Wolf P, Schmidt CS, Meilinger D, Schneider K, Frauer C, et al. Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. J Cell Biochem. 2011;112(9):2585–93.

Prokhortchouk A, Hendrich B, Jørgensen H, Ruzov A, Wilm M, Georgiev G, et al. The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. Genes Dev. 2001;15(13):1613–8.

Prokhortchouk A, Sansom O, Selfridge J, Caballero IM, Salozhin S, Aithozhina D, et al. Kaiso-deficient mice show resistance to intestinal cancer. Mol Cell Biol. 2006;26(1): 199–208.

Quenneville S, Verde G, Corsinotti A, Kapopoulou A, Jakobsson J, Offner S, et al. In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. Mol Cell. 2011;44(3):361–72.

Rais Y, Zviran A, Geula S, Gafni O, Chomsky E, Viukov S, et al. Deterministic direct reprogramming of somatic cells to pluripotency. Nature. 2013;502(7469):65–70.

Rajakumara E, Wang Z, Ma H, Hu L, Chen H, Lin Y, et al. PHD finger recognition of unmodified histone H3R2 links UHRF1 to regulation of euchromatic gene expression. Mol Cell. 2011;43(2):275–84.

Riccio A, Aaltonen LA, Godwin AK, Loukola A, Percesepe A, Salovaara R, et al. The DNA repair gene MBD4 (MED1) is mutated in human carcinomas with microsatellite instability. Nat Genet. 1999;23(3):266–8.

Rothbart SB, Krajewski K, Nady N, Tempel W, Xue S, Badeaux AI, et al. Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. Nat Struct Mol Biol. 2012;19(11):1155–60.

Saito M, Ishikawa F. The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2. J Biol Chem. 2002;277(38):35434–9.

Scarsdale JN, Webb HD, Ginder GD, Williams DC. Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. Nucleic Acids Res. 2011;39(15):6741–52.

Schübeler D. Function and information content of DNA methylation. Nature. 2015;517(7534): 321–6.

Segre JA, Bauer C, Fuchs E. Klf4 is a transcription factor required for establishing the barrier function of the skin. Nat Genet. 1999;22(4):356–60.

Sharif J, Muto M, Takebayashi S-i, Suetake I, Iwamatsu A, Endo TA, et al. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature. 2007;450(7171):908–12.

Shields JM, Yang VW. Identification of the DNA sequence that interacts with the gut-enriched Krüppel-like factor. Nucleic Acids Res. 1998;26(3):796–802.

Skene PJ, Illingworth RS, Webb S, Kerr AR, James KD, Turner DJ, et al. Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. Mol Cell. 2010;37(4):457–68.

Song J, Teplova M, Ishibe-Murakami S, Patel DJ. Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. Science. 2012;335(6069):709–12.

Soubry A, Staes K, Parthoens E, Noppen S, Stove C, Bogaert P, et al. The transcriptional repressor Kaiso localizes at the mitotic spindle and is a constituent of the pericentriolar material. PLoS One. 2010;5(2):e9203.

Stoll R, Lee BM, Debler EW, Laity JH, Wilson IA, Dyson HJ, et al. Structure of the Wilms tumor suppressor protein zinc finger domain bound to DNA. J Mol Biol. 2007;372(5):1227–45.

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2009;324(5929):930–5.

Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006;126(4):663–76.

Talkowski ME, Rosenfeld JA, Blumenthal I, Pillalamarri V, Chiang C, Heilbut A, et al. Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. Cell. 2012;149(3):525–37.

Tate PH, Bird AP. Effects of DNA methylation on DNA-binding proteins and gene expression. Curr Opin Genet Dev. 1993;3(2):226–31.

von Kries JP, Buhrmester H, Strätling WH. A matrix/scaffold attachment region binding protein: identification, purification, and mode of binding. Cell. 1991;64(1):123–35.

Wade PA, Gegonne A, Jones PL, Ballestar E, Aubry F, Wolffe AP. Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. Nat Genet. 1999;23(1):62–6.

Walavalkar NM, Cramer JM, Buchwald WA, Scarsdale JN, Williams DC. Solution structure and intramolecular exchange of methyl-cytosine binding domain protein 4 (MBD4) on DNA suggests a mechanism to scan for mCpG/TpG mismatches. Nucleic Acids Res. 2014;42(17):11218–32.

Weitzel JM, Buhrmester H, Strätling WH. Chicken MAR-binding protein ARBP is homologous to rat methyl-CpG-binding protein MeCP2. Mol Cell Biol. 1997;17(9):5656–66.

Wong E, Yang K, Kuraguchi M, Werling U, Avdievich E, Fan K, et al. Mbd4 inactivation increases C→T transition mutations and promotes gastrointestinal tumor formation. Proc Natl Acad Sci. 2002;99(23):14937–42.

Yildirim O, Li R, Hung J-H, Chen PB, Dong X, Ee L-S, et al. Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. Cell. 2011;147(7):1498–510.

Yoon H-G, Chan DW, Reynolds AB, Qin J, Wong J. N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso. Mol Cell. 2003;12(3):723–34.

Zhang Y, Ng H-H, Erdjument-Bromage H, Tempst P, Bird A, Reinberg D. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev. 1999;13(15):1924–35.

Zhang J, Gao Q, Li P, Liu X, Jia Y, Wu W, et al. S phase-dependent interaction with DNMT1 dictates the role of UHRF1 but not UHRF2 in DNA methylation maintenance. Cell Res. 2011;21(12):1723–39.

Zhao S, Choi M, Overton JD, Bellone S, Roque DM, Cocco E, et al. Landscape of somatic single-nucleotide and copy-number mutations in uterine serous carcinoma. Proc Natl Acad Sci. 2013;110(8):2916–21.

Zhou T, Xiong J, Wang M, Yang N, Wong J, Zhu B, et al. Structural basis for hydroxymethylcytosine recognition by the SRA domain of UHRF2. Mol Cell. 2014;54(5):879–86.

Zuo X, Sheng J, Lau H-T, McDonald CM, Andrade M, Cullen DE, et al. Zinc finger protein ZFP57 requires its co-factor to recruit DNA methyltransferases and maintains DNA methylation imprint in embryonic stem cells via its transcriptional repression domain. J Biol Chem. 2012;287(3):2107–18.

Zviran A, Rais Y, Mor N, Novershtern N, Hanna JH. Mbd3/NuRD is a key inhibitory module during the induction and maintenance of Naïve Pluripotency. bioRxiv. 2015:013961.

DNA Base Flipping: A General Mechanism for Writing, Reading, and Erasing DNA Modifications

Samuel Hong and Xiaodong Cheng

Abstract

The modification of DNA bases is a classic hallmark of epigenetics. Four forms of modified cytosine—5-methylcytosine, 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine—have been discovered in eukaryotic DNA. In addition to cytosine carbon-5 modifications, cytosine and adenine methylated in the exocyclic amine—N4-methylcytosine and N6-methyladenine—are other modified DNA bases discovered even earlier. Each modified base can be considered a distinct epigenetic signal with broader biological implications beyond simple chemical changes. Since 1994, crystal structures of proteins and enzymes involved in writing, reading, and erasing modified bases have become available. Here, we present a structural synopsis of writers, readers, and erasers of the modified bases from prokaryotes and eukaryotes. Despite significant differences in structures and functions, they are remarkably similar regarding their engagement in flipping a target base/nucleotide within DNA for specific recognitions and/or reactions. We thus highlight base flipping as a common structural framework broadly applied by distinct classes of proteins and enzymes across phyla for epigenetic regulations of DNA.

S. Hong (✉)
Department of Biochemistry, Emory University School of Medicine,
1510 Clifton Road, Atlanta, GA 30322, USA

Molecular and Systems Pharmacology Graduate Program,
Emory University School of Medicine, 1510 Clifton Road, Atlanta, GA 30322, USA
e-mail: shong27@emory.edu

X. Cheng
Department of Biochemistry, Emory University School of Medicine,
1510 Clifton Road, Atlanta, GA 30322, USA

© Springer International Publishing Switzerland 2016

A. Jeltsch, R.Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*,
Advances in Experimental Medicine and Biology 945,
DOI 10.1007/978-3-319-43624-1_14

Abbreviations

5caC	5-Carboxylcytosine
5fC	5-Formylcytosine
5ghmC	Glucosylated 5-hydroxymethylcytosine
5hmC	5-Hydroxymethylcytosine
5mC	5-Methylcytosine
AdoHcy	*S*-Adenosyl-L-homocysteine
AdoMet	*S*-Adenosyl-L-methionine
AlkB	*E. coli* alkylated DNA repair protein
ALKBH5	Alkylated DNA repair protein AlkB homolog 5 in human
CMT2	Chromomethylase 2 (plant specific)
CMT3	Chromomethylase 3 (plant specific)
DME	Demeter (plant)
DML3	Demeter-like protein 3 (plant)
DNMT1	Mammalian DNA methyltransferase 1
DNMT3A	Mammalian DNA methyltransferase 3A
DNMT3L	Mammalian DNA methyltransferase 3-like
DRM2	Domain rearranged methyltransferase 2 (plant)
FTO	Fat mass and obesity-associated protein
HhH	Helix-hairpin-helix
JBP	J-binding protein
MBD	Methyl-CpG-binding domain
McrB	Modified cytosine restriction B
Met1	DNA methyltransferase 1 (plant)
MTase	Methyltransferase
N4mC	N4-methylcytosine
N6mA	N6-methyladenine
NOG	N-oxalylglycine
ROS1	Repressor of silencing 1 (plant specific)
SRA	SET and RING associated
TDG	Thymine DNA glycosylase
TET	Ten-eleven translocation
TRD	Target recognition domain
Uhrf1	Ubiquitin-like-containing PHD and RING finger domains protein 1
WH	Winged helix
αKG	α-Ketoglutarate

1 Introduction

Chemical modifications of DNA bases (Fig. 1) have fundamental biological roles in virtually every living organism. In both prokaryotes and many eukaryotes, cytosine can be methylated at the carbon-5 (C5) position by cytosine-C5 methyltransferases

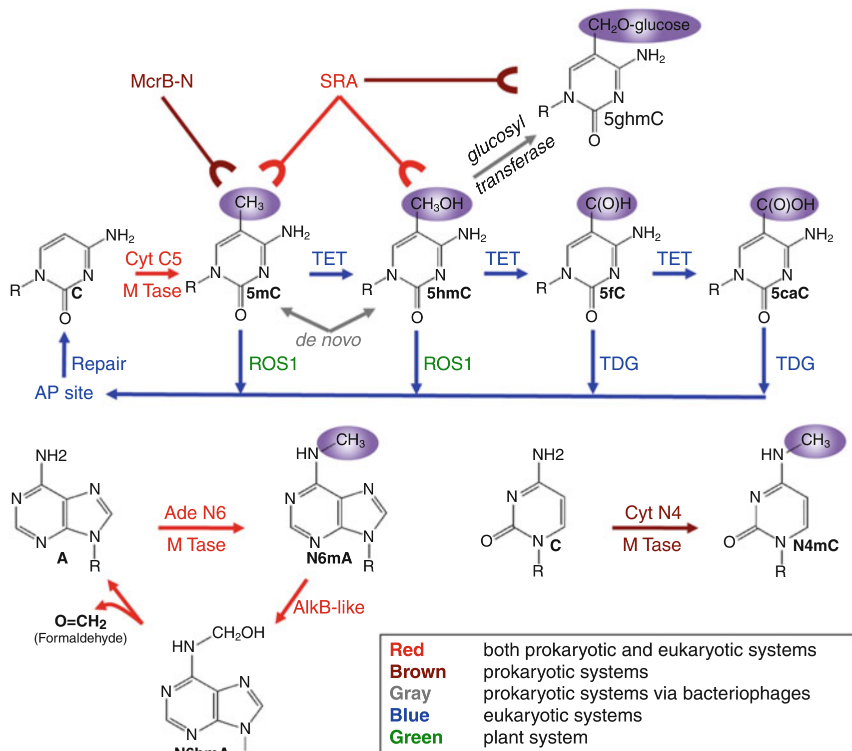


Fig. 1 Chemical modifications of DNA. (**a**) Cytosine-C5 modifications: enzymes and proteins involved in writing, reading, and erasing the modifications via base-flipping mechanisms. (**b**) Adenine-N6 methylation: enzymes involved in writing and erasing DNA adenine N6 methylation. (**c**) Cytosine-N4 methylation

(MTases) to generate 5-methylcytosine (5mC) (Goll and Bestor 2005; Kumar et al. 1994). In higher eukaryotes, 5mC dioxygenases ten-eleven translocation (TET) enzymes utilize α-ketoglutarate (αKG) and Fe(II) to oxidize the methyl group of 5mC to generate 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) via discrete reactions (Ito et al. 2011; Kriaucionis and Heintz 2009; Tahiliani et al. 2009). In prokaryotes, 5mC and 5hmC can be introduced de novo into the genome during phage invasions, as both modified bases can be synthesized prior to incorporation into the phage genome during DNA synthesis (Warren 1980). After DNA synthesis, phage glucosyltransferases can modify 5hmC within the genome to generate glucosylated 5hmC (Kornberg et al. 1961; Lehman and Pratt 1960). Beyond cytosine-C5 modifications, exocyclic amine groups of cytosine and adenine can be methylated in prokaryotes to generate N4-methylcytosine (N4mC) and N6-methyladenine (N6mA) (Cheng 1995; Jeltsch 2002). Crystal structures of DNA modification enzymes to date have consistently shown that the target nucleotide is flipped out of the double helix for reactions in a process called base flipping.

In addition to the modification writers, modified base readers have also been shown to flip the target base for recognitions. Mammalian and plant SET- and RING-associated (SRA) domains recognize 5mC within the genome by base flipping (Arita et al. 2008; Avvakumov et al. 2008; Hashimoto et al. 2008; Rajakumara et al. 2011) and have been characterized as nonenzymatic base flippers. Since the first discovery in eukaryotes, SRA domains have been rediscovered in prokaryotes, recognizing 5mC, 5hmC, and/or 5ghmC to coordinate restriction activity in a modification-dependent manner (Horton et al. 2012, 2014a, b, c). In addition to SRA, the bacterial modified cytosine restriction B enzyme also flips 5mC for recognitions but is structurally distinct from other known base flippers (Sukackaite et al. 2012). Structural homologs of McrB across different phyla may recognize modified bases in a similar way.

A brief survey of DNA base modifications in both prokaryotes and eukaryotes reveals that two major families of enzymes, methyltransferases and dioxygenases, are involved in writing DNA modifications in the four forms of modified cytosine: 5mC, 5hmC, 5fC, and 5caC. In plants, the 5mC DNA glycosylase repressor of silencing 1 (ROS1) can excise 5mC and 5hmC (in vitro) (Gong et al. 2002; Jang et al. 2014; Hong et al. 2014), and in mammals, thymine DNA glycosylase (TDG) can excise 5fC and 5caC (He et al. 2011; Maiti and Drohat 2011; Zhang et al. 2012; Hashimoto et al. 2012a). These discoveries effectively link the base excision repair pathway to DNA demethylation/demodification, by which epigenetic signals encoded in the modified cytosines can be reversed. DNA glycosylases represent the most structurally diverse family of enzymes that are involved in base flipping (also known as nucleotide flipping) (Brooks et al. 2013). Thus, base flipping is not restricted to writers and readers, but has been adopted by DNA glycosylases for erasing DNA modifications as well. Together, structural characterizations of writers, readers, and erasers of DNA base modifications in prokaryotes and eukaryotes effectively showcase base flipping as a general mechanism for regulating and translating fundamental epigenetic signals.

2 Base Flipping for Methylation of DNA Bases

2.1 Bacterial DNMTs (HhaI, TaqI, PvuII)

Biological methylation is widely engaged in various regulations, and it uses *S*-adenosyl-L-methionine (AdoMet) as the primary methyl group donor. The methyl group of AdoMet is bound to a positively charged sulfur atom predisposed to a nucleophilic attack. During the methylation reaction, AdoMet loses the methyl group and becomes *S*-adenosyl-L-homocysteine (AdoHcy). A number of different families of methyltransferases use AdoMet as cofactor, targeting diverse substrates ranging from small molecules to large macromolecules such as DNA, RNA, proteins, lipid, and polysaccharides. The atoms subjected to methylation also vary, including carbon (C), nitrogen (N), oxygen (O), sulfur (S), and several metals. AdoMet-dependent DNA methyltransferases were first discovered in bacterial

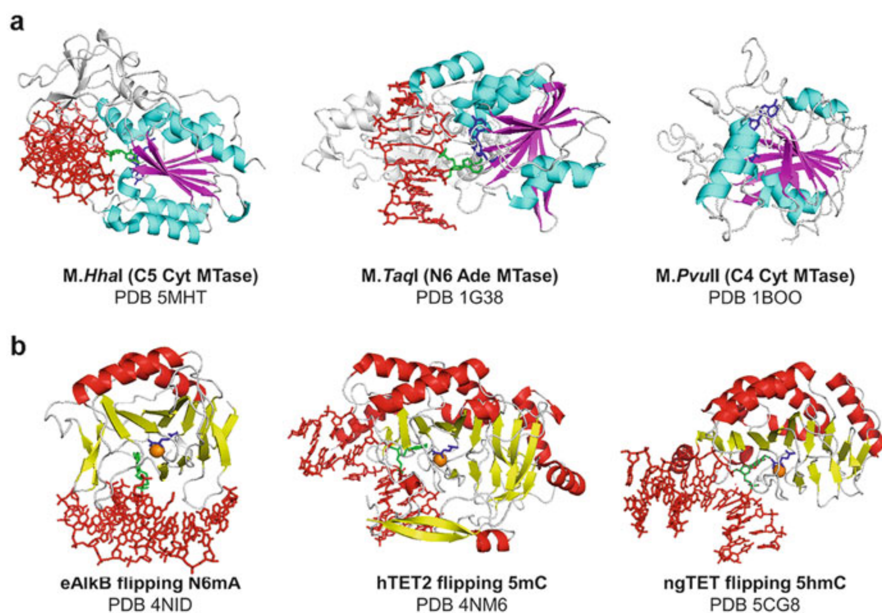


Fig. 2 Writers of DNA modifications. (**a**) Prokaryotic DNA methyltransferases involved in three different types of DNA methylation have similar structures. DNA is colored in *red*, and AdoMet/AdoHcy is colored in *blue*. Flipped bases are shown in *green*. (**b**) *E. coli* AlkB and eukaryotic TET are dioxygenases with common structural folds. αKG is colored in *blue*, and metal in the active sites is colored in *orange*

restriction-modification systems (Roberts et al. 2015). The known structures of AdoMet-dependent DNA methyltransferases share a common "MTase fold" characterized by mixed seven-stranded β sheets (6↓ 7↑ 5↓ 4↓ 1↓ 2↓ 3↓) in which strand 7 is inserted between strands 5 and 6 antiparallel to the others (Cheng and Roberts 2001) (Fig. 2a).

M.HhaI was the first DNA methyltransferase to be structurally characterized (Cheng et al. 1993). It contains an N-terminal MTase domain and a C-terminal target recognition domain (TRD) (Cheng 1995). M.HhaI is a cytosine-C5 methyltransferase that methylates the first cytosine within 5′-GCGC-3′ recognition sequences and prevents R.HhaI restriction activity at the site (Roberts et al. 1976, 2015). Before the structure was available, the proposed mechanism predicted that the catalytic Cys81 would make a nucleophilic attack on the C6 of cytosine to form a covalent complex, followed by transferring the methyl group from AdoMet to cytosine-C5 and releasing the covalent intermediate (Wu and Santi 1985, 1987). In 1994, the crystal structure of M.HhaI-DNA complex with AdoMet was solved as a trapped covalent enzyme-DNA intermediate using 5-fluorocytosine and directly supported the proposed mechanism, presenting the catalytic cysteine covalently linked to C6 and showing methylated C5 adjacent to AdoHcy (Klimasauskas et al. 1994). Yet the most striking aspect of the structure was that both the MTase and the

TRD of the enzyme work simultaneously to bind DNA and flip the target base into the active site pocket. The mechanism of DNA base access by base flipping has since been described as the framework for other DNA methyltransferases (Cheng and Roberts 2001).

After the first structure of M.HhaI-DNA complex was solved, many crystal structures of DNA methyltransferase-DNA complexes have been solved. Besides cytosine-C5 methylation, adenine exocyclic N6 methylation is also a critical modification in prokaryotic DNA (Fig. 1b) and in eukaryotic RNA (Low et al. 2001; Hattman 2005; Jia et al. 2013; Niu et al. 2013). Recent studies have also shown that *Drosophila* genome also harbors N6mA in DNA (Zhang et al. 2015). The structure of the adenine-N6 methyltransferase M.TaqI in complex with DNA and a nonreactive AdoMet analog was solved in 2001 (Goedecke et al. 2001). The enzyme methylates adenine within 5′-TCGA-3′ sequence and harbors a similar two-domain organization as M.HhaI, with the conserved N-terminal MTase domain, but a quite distinct C-terminal TRD (Cheng 1995; Goedecke et al. 2001). The ternary structure is remarkably reminiscent of M.HhaI, involving a flipped adenine in the active site, where the methyl group from the AdoMet analog is positioned near the N6 of the flipped adenine. Instead of the catalytic cysteine residue as in M.HhaI, the asparagine 105 side chain and the following proline backbone oxygen make hydrogen bonds with the adenine-N6 amine group, potentially modulating the direct transfer of the methyl group from AdoMet. A similar mode of interaction is also seen in the active site of the T4 phage DNA adenine methyltransferase (T4 Dam) that flips adenine in 5′-GATC-3′ sequence, and an aspartate residue (Asp171) contacts the adenine-N6 (Horton et al. 2005). Besides adenine-N6 methylation, cytosine-N4 methylation is another type of DNA methylation (Fig. 1c). For example, M.PvuII methylates the central cytosine within 5′-CAGCTG-3′ in the exocyclic amine (Blumenthal et al. 1985; Bheemanaik et al. 2006). The structure of M.PvuII is available only in an AdoMet-bound form without DNA, yet it contains many shared features of other methyltransferases in terms of domain organization and AdoMet interactions (Gong et al. 1997; Bheemanaik et al. 2006).

2.2 Mammalian DNMTs (DNMT1, DNMT3A/3L)

Structural features of classic prokaryotic methyltransferases are extensively shared by the mammalian DNA methyltransferases DNMT1, DNMT3A, and DNMT3B. They are all cytosine-C5 methyltransferases containing an MTase domain with a catalytic cysteine and a TRD. DNMT1 is primarily implicated in methylation of the daughter strand during DNA replication to maintain the methylation pattern encoded in the mother strand by preferentially recognizing hemimethylated DNA in CpG dinucleotide context (Li et al. 1992; Yoder et al. 1997). On the other hand, DNMT3A and DNMT3B are considered de novo methyltransferases that can methylate CpG sites as well as non-CpG sites (Ramsahoye et al. 2000; Gowher and Jeltsch 2001; Suetake et al. 2003). Such differences in substrate specificities are partly due to the involvement of other domains outside the catalytic fragment. For example, a CXXC domain and a BAH1 domain within DNMT1 hinder

methylation of unmethylated CpG sites (Song et al. 2011), whereas DNMT3A and DNMT3B do not contain such domains and can readily methylate them.

2.3 Implications of DNA Methyltransferase Dimers (DNMT3A/3L and EcoP15I)

Besides being a catalytic domain, the MTase domain can participate in protein-protein interactions as exemplified by the DNMT3A MTase domain interacting with a naturally inactive MTase-like domain of DNMT3L, a scaffold protein that binds histone tail H3 to guide DNMT3A activities by forming a tetramer of 3L-3A-3A-3L (Jia et al. 2007; Ooi et al. 2007). Interestingly, a multi-subunit prokaryotic DNA N6mA methyltransferase, EcoP15I, contains a DNA MTase dimer in which one monomer is involved in target base flipping and the other in the recognition of DNA base context (Gupta et al. 2015). Thus, dimerization of two structurally comparable proteins for divergent functionalities may be a mechanism for fine-controlling genomic DNA modifications.

2.4 Plant DNMTs

Plant DNA MTases show similar functionalities as the mammalian counterparts. Met1 is homologous to mammalian DNMT1 and is responsible for the maintenance of CpG methylation, whereas domain rearranged methyltransferase 2 (DRM2) is involved in de novo DNA methylation (Goll and Bestor 2005; Law and Jacobsen 2010). DRM2 contains a rearranged MTase domain, such that its N-terminal half is equivalent to the C-terminal half of the conventional MTase fold and vice versa. A structural study of DRM2 family MTase domain has revealed that the rearranged domain still forms a classic MTase structure and functions as a homodimer (Zhong et al. 2014) analogous to DNMT3A-3L heterodimer. In addition to Met1 and DRM2, plants also have plant-specific DNA methyltransferases, such as CMT2 and CMT3 that are specifically involved in non-CpG methylation (Stroud et al. 2014; Lindroth et al. 2001; Zemach et al. 2013). The higher diversity of the MTase family within plants compared to the mammalian family suggests that DNA methylation may be more dynamically regulated in plants than in mammals.

3 Base Flipping in Oxidative Modifications of Methylated Bases

3.1 Eukaryotic TET Enzymes

The 5mC is by far the most widely studied modified base. Yet, if 5mC has been considered “the fifth” base, 5hmC is increasingly being labeled as “the sixth” base and has garnered much attention recently. The existence of 5hmC in bacteriophage, modified from 2′-deoxycytidine before the integration into the viral genome (Warren

1980), was first reported in 1953 (Wyatt and Cohen 1953). In 1993, a novel J base (β-D-glucosyl-hydroxymethyluracil) was discovered in trypanosomes, in which J-binding proteins (JBP1 and JBP2) are involved in oxidizing the C5-methyl group of thymine during J-base synthesis by using αKG and Fe(II) as cofactors to generate 5-hydroxymethyluracil (Gommers-Ampt et al. 1993; Borst and Sabatini 2008). In 2009, mammalian JBP homolog TET enzymes were discovered to oxidize the methyl group of 5mC to generate 5hmC (Tahiliani et al. 2009; Kriaucionis and Heintz 2009). Further analysis revealed that TET enzymes could further oxidize 5hmC to 5fC and then to 5caC (Ito et al. 2011). Also, TET enzymes have been shown to convert thymine (5-methyluracil) to 5-hydroxymethyluracil by oxidizing the C5-methyl group of thymine (Pfaffeneder et al. 2014; Pais et al. 2015).

Eukaryotic JBP/TET homologs are present across many eukaryotic organisms including amoeboflagellate *Naegleria gruberi* (Iyer et al. 2013; Hashimoto et al. 2014b). Crystal structures of *Naegleria gruberi* TET-like (NgTET) and human TET2 (hTET2) in complex with 5mC-, 5hmC-, and 5fC-containing DNA have been characterized (Hashimoto et al. 2014b, 2015a; Hu et al. 2013a, 2015). All TET structures show a flipped base positioned in the active site pocket close to N-oxalylglycine (NOG)—an inactive αKG analog—and a divalent metal such as Fe(II) or Mn(II) used for stalling the enzyme in the pre-reaction state. Some of the features of flipped base recognitions observed in DNMT-DNA complex structures (Cheng and Roberts 2001; Horton et al. 2005) can also be seen in the structures of TET-DNA complexes. The flipped base in the active site of a TET enzyme in complex with DNA is stabilized by π stacking interactions involving an aromatic residue such as Phe295 in NgTET (Hashimoto et al. 2014b) and Tyr1902 in hTET2 (Hu et al. 2013a). Also, polar residues such as Asn147, His297, and Asp234 in NgTET contact O2, N3, and N4, respectively, to guide substrate specificities (Hashimoto et al. 2014b), and the methyl or the hydroxymethyl group is oriented toward NOG and Fe(II)/Mn(II) (Hashimoto et al. 2015a; Hu et al. 2015). Often, active site pockets for flipped bases not only contain residues for base recognition but also specifically orient the base for distinct reactions depending on the type of enzymes. Base flipping is therefore a common mechanism applied by different classes of enzymes, such as AdoMet-dependent methyltransferases and αKG- and Fe(II)-dependent dioxygenases to recognize and stabilize the target base for specific reactions.

3.2 AlkB and Homologs

Similar to TET enzymes, eukaryotic homologs of *E. coli* AlkB such as FTO and ALKBH5 are also αKG- and Fe(II)-dependent dioxygenases that can oxidize the methyl group of N6mA within mRNA to yield demethylated adenine (Jia et al. 2011; Zheng et al. 2013; Zhu and Yi 2014). Indeed, TET-DNA complex structures are remarkably comparable to that of the AlkB-DNA complex, and both TET enzymes and AlkB homologs perform base flipping as part of their reaction mechanism (Hu et al. 2013a; Iyer et al. 2013; Hashimoto et al. 2014b; Zhu and Yi 2014; McDonough et al. 2010). Common structural folds include two twisted β-sheets in

the core where the active site is formed (Fig. 2b). However, the two enzyme families differ in an important way. TET enzymes oxidize CH_3 attached to an inert carbon atom (cytosine or thymine C5). The resulting product 5hmC (or 5hmU) is very stable and can undergo further oxidations in subsequent rounds of reactions to generate further oxidized products. On the other hand, FTO and ALKBH5 likely generate N6-hydroxymethyladenine intermediate in which the oxidized carbon is attached to a reactive nitrogen atom (adenine-N6). This intermediate spontaneously releases the hydroxymethyl group as formaldehyde and decomposes to adenine—the final "demethylated" product (Hashimoto et al. 2015b) (Fig. 1b). Therefore, AlkB and its homologs are demethylases, while TET enzymes should not be designated as a demethylase but would rather be appropriately understood as a "writer" that generates additional marks on 5mC within genomes to alter epigenetic signals.

Several biochemical observations suggest that modified cytosines beyond 5mC may form distinct epigenetic signals. Many 5mCpG readers such as methyl-CpG binding domain (MBD) proteins have shown significantly reduced binding affinity toward 5hmC when compared to 5mC within CpG context (Hashimoto et al. 2012b), whereas some proteins may preferentially bind 5hmC (Spruijt et al. 2013). DNMT1 has a significantly reduced affinity toward hemi-hydroxymethylated DNA substrate compared to hemi-methylated DNA (Hashimoto et al. 2012b), suggesting that methylation marks altered by TET enzymes can be lost in subsequent DNA replications. In addition, the RNA polymerase II transcription rate can be specifically reduced by 5fC and 5caC (Kellinger et al. 2012; Wang et al. 2015). These findings strongly point to the possibility that modifications beyond 5mC are distinct signals, and much future work is needed to elucidate how the modified bases are differently implicated in larger biological contexts.

4 Base Flipping in the Recognition of Modified Bases

4.1 Eukaryotic SRA Domains

The function of 5mC and N6mA in prokaryotes was classically understood in the context of restriction-modification systems, in which methylated bacterial DNA is protected from restriction digestion (Wilson and Murray 1991). Effects of DNA methylation are fundamentally determined by the way the methyl groups alter various protein-DNA interactions. In eukaryotes, genomic 5mC bases are widely involved in various regulatory processes to control gene expression, chromatin states, and genomic stability that are highly relevant in the human disease context (Robertson 2005). Such penetrating biological implications can be partly attributed to a large number of protein-DNA interactions that are potentially affected by DNA methylation in a direct manner. Evidence shows that several transcription factors are prevented from DNA binding when the binding site is methylated (Tate and Bird 1993), whereas several MBD family proteins are specific 5mCpG readers, as previously mentioned (Klose and Bird 2006). The interface between methylated DNA and its biological effects can be further complicated by the involvement of the

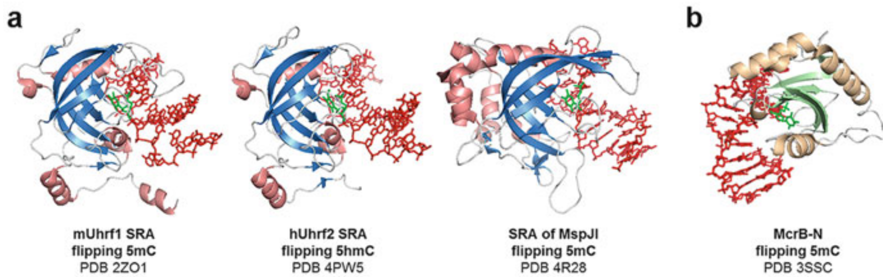


Fig. 3 Readers of DNA modifications. (**a**) Prokaryotic and eukaryotic SRA domains recognize C5-modified cytosine via base flipping and are similar in structures. (**b**) Crystal structure of prokaryotic McrB-N monomer flipping 5mC

nucleosome context which is closely interwoven with DNA methylation (Hashimoto et al. 2010; Cedar and Bergman 2009).

The initial discovery of 5mC-binding proteins has raised the possibility of other readers involved in modified base recognitions. In 2007, another family of 5mC readers was discovered in plants and was termed SET and RING associated (SRA) domain as a part of VIM1 (Woo et al. 2007). A mammalian homolog to VIM1 is Uhrf1, which can associate with DNMT1 during post-replicative maintenance of DNA methylation (Bostick et al. 2007). In the following year, three crystal structures of the mammalian Uhrf1 SRA domain in complex with 5mC-containing DNA were reported (Hashimoto et al. 2008; Avvakumov et al. 2008; Arita et al. 2008). The structures have revealed that SRA recognizes 5mC by base flipping, although it is not a DNA-modifying enzyme such as methyltransferases or dioxygenases. SRA is also structurally distinct from other base flippers and is characterized by a twisted β-sheet fold resembling a half-moon shape (Fig. 3a). Remarkably, the 5mC-binding pocket of SRA features familiar modes of base recognitions exemplified by π stacking interactions, recognitions of the N3 and N4 by Asp474 side chain, and a van der Waals contact of the C5-methyl group of flipped 5mC by Ser486 Cβ.

Interestingly, the SRA of Uhrf2 binds 5hmC with a slightly higher preference compared to 5mC, and the crystal structure of Uhrf2 SRA in complex with 5hmC-containing DNA is available (Zhou et al. 2014). In the structure, 5hmC is flipped and stabilized, and the OH moiety of the hydroxymethyl group is contacted by the backbone carbonyl groups of Thr508 and Gly509 in the active site pocket which is slightly larger in size compared to that of Uhrf1 SRA. Therefore, the eukaryotic SRA domain has been characterized as a base-flipping domain that recognizes both 5mC and 5hmC.

4.2 Prokaryotic SRA Domains

Recently, SRA domains have been rediscovered in prokaryotes in families of modification-dependent restriction enzymes that recognize modified bases and introduce a double-stranded break in some distances away. MspJI was among the first such enzymes to be reported, which recognizes hemi-modified 5mC or 5hmC by the N-terminal SRA-like domain and restricts the DNA by the C-terminal

endonuclease domain (Cohen-Karni et al. 2011). The crystal structure of MspJI has been solved with substrate DNA, revealing an SRA-like structure in the N-terminal modification recognition domain that flips the target 5mC (Cohen-Karni et al. 2011; Horton et al. 2014c). Despite the lack of amino acid sequence conservation between eukaryotic UHRF1/2 SRA and MspJI SRA, all SRA domains feature a twisted β-sheet fold with a half-moon shape (Fig. 3a).

As more modification-dependent restriction enzymes have been identified, some of them are found with different specificities toward 5mC, 5hmC, and 5ghmC. AbaSI, unlike MspJI, has an N-terminal Vsr-like endonuclease domain and a C-terminal SRA-like domain (Borgaro and Zhu 2013; Horton et al. 2014a). Its SRA domain seems to preferentially recognize 5ghmC and 5hmC compared to 5mC, as the relative rate of cleavage of DNA containing the corresponding modification is 5ghmC:5hmC:5mC = 8000:500:1 (Wang et al. 2011). Structural features within SRA domains that fine-tune such specificities await future characterizations.

4.3 McrB-N as Distinct 5mC Reader

Modification-dependent restriction enzymes also utilize yet another 5mC recognition domain (Fig. 3b). The N-terminus of McrB (McrB-N) recognizes 5mC next to adenine within 5′-ACCGGT-3′ sequences, and McrC associates with McrB to provide endonuclease activity (Sutherland et al. 1992; Gast et al. 1997). The crystal structure of McrB-N in complex with 5mC-containing DNA shows flipped 5mC in the active site, revealing a novel fold distinct from any other known base flippers (Sukackaite et al. 2012). The active site displays familiar π stacking of the flipped 5mC via aromatic residues and van der Waals contact of the C5-methyl group via the side chain of Leu68. So far, SRA is the only known modified base reader in eukaryotes that flips the target base, and no eukaryotic homolog of McrB-N has been identified. However, the history of the discovery of base flippers suggests a strong possibility of its structural homologs present in a wide spectrum of phyla.

4.4 DpnI as N6mA Reader

While base flipping seems to be a major mechanism by which a modified DNA base can be recognized, it should be noted that modified bases can be recognized by some transcription factors in a sequence-dependent context as well (Spruijt et al. 2013; Hu et al. 2013b), none of which involves base flipping. Along with the previously mentioned MBD family proteins that recognize 5mC within the simple dinucleotide CpG sequence, certain mammalian zinc-finger family proteins such as Kaiso (Buck-Koehntop et al. 2012), Zfp57 (Liu et al. 2012), Klf4 (Liu et al. 2014), and Egr1 (Hashimoto et al. 2014a) bind 5mC within specific sequences via a common structural motif (Liu et al. 2013; Hashimoto et al. 2015b). In addition, another zinc-finger transcription factor WT1 (Hashimoto et al. 2014a) and the basic helix-loop-helix (bHLH) family Tcf3-Ascl1 heterodimer (Golla et al. 2014) can specifically bind 5caC within their consensus sequences. In prokaryotes, DpnI

harbors a C-terminal winged-helix (WH) domain that recognizes the methyl group of N6mA within 5′-GATC-3′ sequence via Trp138 involving van der Waals interactions (Mierzejewska et al. 2014). Therefore, DNA modifications may regulate transcription-binding sites in much more dynamic and selective manners than they were previously understood.

5 Base Flipping in Removing Modified and Unmodified Bases

5.1 Mammalian Thymine DNA Glycosylase (TDG)

The discovery of TET-mediated modified cytosine bases has provided a fresh insight into a long sought-after pathway of 5mC demethylation/demodification within mammalian genomes (see review (Zhu 2009)). In the base excision repair pathway, DNA glycosylases cleave the glycosidic bond between the ribose and the target base and represent the most structurally diverse family of base-flipping enzymes (Brooks et al. 2013). Initially, it was hypothesized that 5mC is removed by 5mC DNA glycosylase(s), as mammalian 5mC DNA glycosylase activity had been reported (Vairapandi and Duker 1993, 1996; Vairapandi et al. 2000). However, the glycosylase involved was never identified. After the discovery of TET enzymes, mammalian TDG that generally removes uracil or thymine mismatched to guanine was surprisingly revealed to excise 5fC and 5caC to establish genome-wide DNA demethylation (He et al. 2011; Maiti and Drohat 2011; Hashimoto et al. 2012a; Zhang et al. 2012). The crystal structure of the human TDG catalytic domain in complex with 5caC-containing DNA was also solved (Fig. 4a), presenting the flipped base in the active site where the C5-carboxyl moiety of 5caC is specifically recognized by the

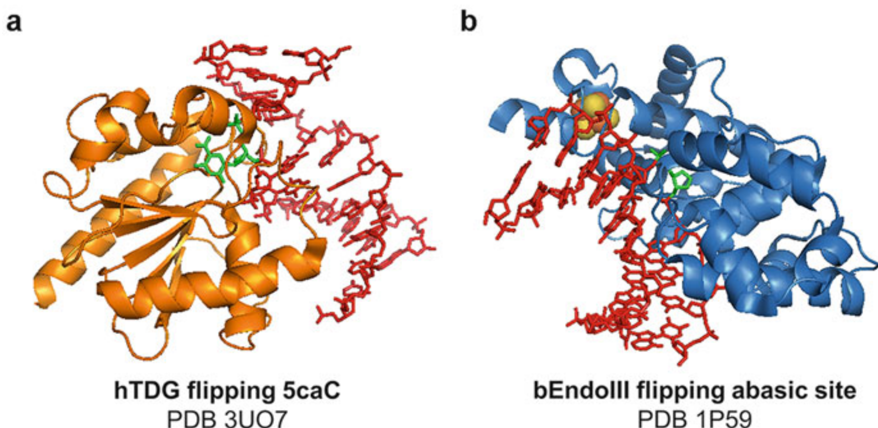


Fig. 4 Erasers of DNA modifications. (**a**) Crystal structure of human TDG flipping 5caC opposite guanine. (**b**) Crystal structure of *Geobacillus stearothermophilus* endonuclease III in complex with DNA. Iron-sulfur cluster is colored in *orange* and *yellow*

side chain of Asn157 and the Tyr152 amide backbone (Zhang et al. 2012). The discovery of TDG excising 5fC and 5caC has effectively linked the base excision repair pathway to DNA demethylation in mammalian system.

5.2 Plant ROS1

In plants, paradoxically, bona fide 5mC DNA glycosylases were clearly demonstrated and identified in 2002 (Gong et al. 2002), approximately a decade before TET and TDG were implicated in DNA demethylation. In *Arabidopsis*, four closely related 5mC DNA glycosylases exist: ROS1, DME, DML2, and DML3 (Gong et al. 2002; Morales-Ruiz et al. 2006; Gehring et al. 2006; Ortega-Galisteo et al. 2008). They have a catalytic glycosylase domain homologous to *E. coli* endonuclease III (Fig. 4b), a helix-hairpin-helix (HhH) fold DNA glycosylase that harbors an iron-sulfur cluster-binding site and excises damaged pyrimidines (Ponferrada-Marin et al. 2009, 2011; Mok et al. 2010). Both ROS1 and DME have been shown to excise 5mC in vivo and in vitro (Gong et al. 2002; Ponferrada-Marin et al. 2009; Gehring et al. 2006; Mok et al. 2010), and they are shown to excise 5hmC, but not 5fC and 5caC in vitro (Hong et al. 2014; Jang et al. 2014; Brooks et al. 2014). Thus, plant ROS1 and mammalian TDG have mutually exclusive substrate specificities for 5mC, 5hmC, 5fC, and 5caC; the first two are substrates for ROS1 and the latter two for TDG (Hashimoto et al. 2012a; Hong et al. 2014). One of the most surprising aspects of plant 5mC DNA glycosylases is that they excise the target base only when both the catalytic glycosylase domain and the C-terminal domain are present (Hong et al. 2014; Mok et al. 2010). The C-terminal domain of ROS1 is conserved only among plant 5mC DNA glycosylases and has been shown to strongly associate with the catalytic domain, suggesting that domain-domain interactions are important for target base recognition and excision (Hong et al. 2014).

While TDG and ROS1 have been clearly implicated in DNA demethylation pathways, jury is still out on the possibility of the contribution of other pathways to DNA demethylation. In addition to the previously mentioned mammalian 5mC DNA glycosylase activities, 5hmC DNA glycosylase activity was observed in a calf thymus extract (Cannon et al. 1988). A recent proteomic study has revealed that several mammalian DNA glycosylases such as NTH1, OGG1, NEIL1, and NEIL2 bind 5mC- and 5hmC-containing DNA in a modification-specific manner (Spruijt et al. 2013), though they by themselves do not have the glycosylase activity against 5mC or 5hmC (Hong et al. 2014).

The 5mC DNA glycosylase activity by ROS1 is interesting from a standpoint of historical characterization of DNA glycosylases as DNA damage repair enzymes. In a given genome, there can be many types of damaged bases, and their diversity is on par with many classes of DNA glycosylases that are structurally distinct (Brooks et al. 2013). On the other hand, 5mC in plants is not considered a damaged base and exists in substantial amounts in the *Arabidopsis* genome (Zhang et al. 2006). Thus, ROS1 must be regulated and specifically targeted to a certain genomic

location to initiate DNA demethylation (Zheng et al. 2008; Qian et al. 2012). In addition to 5mC, ROS1 is comparably active on thymine mismatched to guanine and on some damaged pyrimidines, suggesting that ROS1 can be involved in both DNA demethylation and DNA damage repair (Ponferrada-Marin et al. 2009, 2010). Such dual functionality can be applied to TDG as well, which not only excises thymine or uracil mismatched to G during the process of DNA mismatch repair but also excises 5fC and 5caC base paired with guanine for DNA demodification in mammals (He et al. 2011; Maiti and Drohat 2011; Hashimoto et al. 2012a; Zhang et al. 2012).

5.3 Achaeon PabI Activity as Adenine DNA Glycosylase

Interestingly, the archaeal *Pyrococcus abyssi* PabI enzyme was initially thought to be a restriction endonuclease but has recently been re-characterized as a sequence-specific adenine DNA glycosylase (Miyazono et al. 2014). PabI is comparable to MutY family mismatch repair DNA glycosylases that excise target adenine mismatched to 8-oxoguanine (Fromme et al. 2004). However PabI is remarkably distinct from MutY, because PabI excises adenine correctly base paired to thymine in a targeted manner. It is therefore possible that DNA glycosylases have adapted to function in more processes than DNA damage repair by removing benign bases for various biological regulations.

Conclusions

First observed in 1994 in the crystal structure of M.HhaI with DNA, base flipping is now understood as a common mode of protein-DNA/RNA interactions adopted by structurally and functionally distinct classes of proteins across various phyla. Base flipping is the only known mechanism for establishing DNA modifications in a targeted manner via DNA methyltransferases and TET dioxygenases. What used to be considered a eukaryote-specific base-flipping 5mC reader, SRA, has later been shown to be widely prevalent in prokaryotic systems for recognizing several modified bases including 5mC, 5hmC, and 5ghmC. In addition to SRA, more structurally diverse classes of modified base readers have been discovered in prokaryotes, such as the base-flipping 5mC reader McrB-N and the N6mA-recognizing WH domain of DpnI (using non-base-flipping mechanism). Also, DNA glycosylases are base flippers primarily characterized as DNA repair enzymes, though not all DNA glycosylases flip a base/nucleotide for base excision, as presented in the very recent example of bacterial AlkD (Mullins et al. 2015). Today, DNA demodification is considered a bona fide output of the base excision repair pathway through DNA glycosylases, such as mammalian TDG and plant ROS1 whose mechanism of action again involves base flipping. In an era in which DNA modifications are considered critical and increasingly complex epigenetic signals, this simple, but elegant, structural mechanism for protein-DNA interaction is preserved as a truly ubiquitous framework.

Acknowledgments The work in the authors' laboratory is supported by grant from National Institutes of Health (GM049245-22). X.C. is a Georgia Research Alliance Eminent Scholar.

References

Arita K, Ariyoshi M, Tochio H, Nakamura Y, Shirakawa M. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature. 2008;455(7214):818–21. doi:10.1038/nature07249.

Avvakumov GV, Walker JR, Xue S, Li Y, Duan S, Bronner C, et al. Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. Nature. 2008;455(7214):822–5. doi:10.1038/nature07273.

Bheemanaik S, Reddy YV, Rao DN. Structure, function and mechanism of exocyclic DNA methyltransferases. Biochem J. 2006;399(2):177–90. doi:10.1042/BJ20060854.

Blumenthal RM, Gregory SA, Cooperider JS. Cloning of a restriction-modification system from Proteus vulgaris and its use in analyzing a methylase-sensitive phenotype in Escherichia coli. J Bacteriol. 1985;164(2):501–9.

Borgaro JG, Zhu Z. Characterization of the 5-hydroxymethylcytosine-specific DNA restriction endonucleases. Nucleic Acids Res. 2013;41(7):4198–206. doi:10.1093/nar/gkt102.

Borst P, Sabatini R. Base J: discovery, biosynthesis, and possible functions. Annu Rev Microbiol. 2008;62:235–51. doi:10.1146/annurev.micro.62.081307.162750.

Bostick M, Kim JK, Esteve PO, Clark A, Pradhan S, Jacobsen SE. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science. 2007;317(5845):1760–4. doi:10.1126/science.1147939.

Brooks SC, Adhikary S, Rubinson EH, Eichman BF. Recent advances in the structural mechanisms of DNA glycosylases. Biochim Biophys Acta. 2013;1834(1):247–71. doi:10.1016/j.bbapap.2012.10.005.

Brooks SC, Fischer RL, Huh JH, Eichman BF. 5-methylcytosine recognition by Arabidopsis thaliana DNA glycosylases DEMETER and DML3. Biochemistry. 2014;53(15):2525–32. doi:10.1021/bi5002294.

Buck-Koehntop BA, Stanfield RL, Ekiert DC, Martinez-Yamout MA, Dyson HJ, Wilson IA, et al. Molecular basis for recognition of methylated and specific DNA sequences by the zinc finger protein Kaiso. Proc Natl Acad Sci U S A. 2012;109(38):15229–34. doi:10.1073/pnas.1213726109.

Cannon SV, Cummings A, Teebor GW. 5-Hydroxymethylcytosine DNA glycosylase activity in mammalian tissue. Biochem Biophys Res Commun. 1988;151(3):1173–9.

Cedar H, Bergman Y. Linking DNA methylation and histone modification: patterns and paradigms. Nat Rev Genet. 2009;10(5):295–304. doi:10.1038/nrg2540.

Cheng X. Structure and function of DNA methyltransferases. Annu Rev Biophys Biomol Struct. 1995;24:293–318. doi:10.1146/annurev.bb.24.060195.001453.

Cheng X, Roberts RJ. AdoMet-dependent methylation, DNA methyltransferases and base flipping. Nucleic Acids Res. 2001;29(18):3784–95.

Cheng X, Kumar S, Posfai J, Pflugrath JW, Roberts RJ. Crystal structure of the HhaI DNA methyltransferase complexed with S-adenosyl-L-methionine. Cell. 1993;74(2):299–307.

Cohen-Karni D, Xu D, Apone L, Fomenkov A, Sun Z, Davis PJ, et al. The MspJI family of modification-dependent restriction endonucleases for epigenetic studies. Proc Natl Acad Sci U S A. 2011;108(27):11040–5. doi:10.1073/pnas.1018448108.

Fromme JC, Banerjee A, Huang SJ, Verdine GL. Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase. Nature. 2004;427(6975):652–6. doi:10.1038/nature02306.

Gast FU, Brinkmann T, Pieper U, Kruger T, Noyer-Weidner M, Pingoud A. The recognition of methylated DNA by the GTP-dependent restriction endonuclease McrBC resides in the N-terminal domain of McrB. Biol Chem. 1997;378(9):975–82.

Gehring M, Huh JH, Hsieh TF, Penterman J, Choi Y, Harada JJ, et al. DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. Cell. 2006;124(3):495–506. doi:10.1016/j.cell.2005.12.034.

Goedecke K, Pignot M, Goody RS, Scheidig AJ, Weinhold E. Structure of the N6-adenine DNA methyltransferase M.TaqI in complex with DNA and a cofactor analog. Nat Struct Biol. 2001;8(2):121–5. doi:10.1038/84104.

Goll MG, Bestor TH. Eukaryotic cytosine methyltransferases. Annu Rev Biochem. 2005;74:481–514. doi:10.1146/annurev.biochem.74.010904.153721.

Golla JP, Zhao J, Mann IK, Sayeed SK, Mandal A, Rose RB, et al. Carboxylation of cytosine (5caC) in the CG dinucleotide in the E-box motif (CGCAG|GTG) increases binding of the Tcf3|Ascl1 helix-loop-helix heterodimer 10-fold. Biochem Biophys Res Commun. 2014;449(2):248–55. doi:10.1016/j.bbrc.2014.05.018.

Gommers-Ampt JH, Van Leeuwen F, de Beer AL, Vliegenthart JF, Dizdaroglu M, Kowalak JA, et al. beta-D-glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan T. brucei. Cell. 1993;75(6):1129–36.

Gong W, O'Gara M, Blumenthal RM, Cheng X. Structure of pvu II DNA-(cytosine N4) methyltransferase, an example of domain permutation and protein fold assignment. Nucleic Acids Res. 1997;25(14):2702–15.

Gong Z, Morales-Ruiz T, Ariza RR, Roldan-Arjona T, David L, Zhu JK. ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. Cell. 2002;111(6):803–14.

Gowher H, Jeltsch A. Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpG [correction of non-CpA] sites. J Mol Biol. 2001;309(5):1201–8. doi:10.1006/jmbi.2001.4710.

Gupta YK, Chan SH, Xu SY, Aggarwal AK. Structural basis of asymmetric DNA methylation and ATP-triggered long-range diffusion by EcoP15I. Nat Commun. 2015;6:7363. doi:10.1038/ncomms8363.

Hashimoto H, Horton JR, Zhang X, Bostick M, Jacobsen SE, Cheng X. The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. Nature. 2008;455(7214):826–9. doi:10.1038/nature07280.

Hashimoto H, Vertino PM, Cheng X. Molecular coupling of DNA methylation and histone methylation. Epigenomics. 2010;2(5):657–69. doi:10.2217/epi.10.44.

Hashimoto H, Hong S, Bhagwat AS, Zhang X, Cheng X. Excision of 5-hydroxymethyluracil and 5-carboxylcytosine by the thymine DNA glycosylase domain: its structural basis and implications for active DNA demethylation. Nucleic Acids Res. 2012a;40(20):10203–14. doi:10.1093/nar/gks845.

Hashimoto H, Liu Y, Upadhyay AK, Chang Y, Howerton SB, Vertino PM, et al. Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Res. 2012b;40(11):4841–9. doi:10.1093/nar/gks155.

Hashimoto H, Olanrewaju YO, Zheng Y, Wilson GG, Zhang X, Cheng X. Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence. Genes Dev. 2014a;28(20):2304–13. doi:10.1101/gad.250746.114.

Hashimoto H, Pais JE, Zhang X, Saleh L, Fu ZQ, Dai N, et al. Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. Nature. 2014b;506(7488):391–5. doi:10.1038/nature12905.

Hashimoto H, Pais JE, Dai N, Correa Jr IR, Zhang X, Zheng Y, et al. Structure of Naegleria Tet-like dioxygenase (NgTet1) in complexes with a reaction intermediate 5-hydroxymethylcytosine DNA. Nucleic Acids Res. 2015a;43(22):10713–21. doi:10.1093/nar/gkv870.

Hashimoto H, Zhang X, Vertino PM, Cheng X. The mechanisms of generation, recognition, and erasure of DNA 5-methylcytosine and thymine oxidations. J Biol Chem. 2015b;290(34):20723–33. doi:10.1074/jbc.R115.656884.

Hattman S. DNA-[adenine] methylation in lower eukaryotes. Biochem Biokhimiia. 2005;70(5):550–8.

He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science. 2011;333(6047):1303–7. doi:10.1126/science.1210944.

Hong S, Hashimoto H, Kow YW, Zhang X, Cheng X. The carboxy-terminal domain of ROS1 is essential for 5-methylcytosine DNA glycosylase activity. J Mol Biol. 2014;426(22):3703–12. doi:10.1016/j.jmb.2014.09.010.

Horton JR, Liebert K, Hattman S, Jeltsch A, Cheng X. Transition from nonspecific to specific DNA interactions along the substrate-recognition pathway of dam methyltransferase. Cell. 2005;121(3):349–61. doi:10.1016/j.cell.2005.02.021.

Horton JR, Mabuchi MY, Cohen-Karni D, Zhang X, Griggs RM, Samaranayake M, et al. Structure and cleavage activity of the tetrameric MspJI DNA modification-dependent restriction endonuclease. Nucleic Acids Res. 2012;40(19):9763–73. doi:10.1093/nar/gks719.

Horton JR, Borgaro JG, Griggs RM, Quimby A, Guan S, Zhang X, et al. Structure of 5-hydroxy methylcytosine-specific restriction enzyme, AbaSI, in complex with DNA. Nucleic Acids Res. 2014a;42(12):7947–59. doi:10.1093/nar/gku497.

Horton JR, Nugent RL, Li A, Mabuchi MY, Fomenkov A, Cohen-Karni D, et al. Structure and mutagenesis of the DNA modification-dependent restriction endonuclease AspBHI. Sci Rep. 2014b;4:4246. doi:10.1038/srep04246.

Horton JR, Wang H, Mabuchi MY, Zhang X, Roberts RJ, Zheng Y, et al. Modification-dependent restriction endonuclease, MspJI, flips 5-methylcytosine out of the DNA helix. Nucleic Acids Res. 2014c;42(19):12092–101. doi:10.1093/nar/gku871.

Hu L, Li Z, Cheng J, Rao Q, Gong W, Liu M, et al. Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. Cell. 2013a;155(7):1545–55. doi:10.1016/j.cell.2013.11.020.

Hu S, Wan J, Su Y, Song Q, Zeng Y, Nguyen HN, et al. DNA methylation presents distinct binding sites for human transcription factors. Elife. 2013b;2:e00726. doi:10.7554/eLife.00726.

Hu L, Lu J, Cheng J, Rao Q, Li Z, Hou H, et al. Structural insight into substrate preference for TET-mediated oxidation. Nature. 2015;527(7576):118–22. doi:10.1038/nature15713.

Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science. 2011;333(6047):1300–3. doi:10.1126/science.1210597.

Iyer LM, Zhang D, Burroughs AM, Aravind L. Computational identification of novel biochemical systems involved in oxidation, glycosylation and other complex modifications of bases in DNA. Nucleic Acids Res. 2013;41(16):7635–55. doi:10.1093/nar/gkt573.

Jang H, Shin H, Eichman BF, Huh JH. Excision of 5-hydroxymethylcytosine by DEMETER family DNA glycosylases. Biochem Biophys Res Commun. 2014;446(4):1067–72. doi:10.1016/j.bbrc.2014.03.060.

Jeltsch A. Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. Chembiochem Eur J chem biol. 2002;3(4):274–93.

Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature. 2007;449(7159):248–51. doi:10.1038/nature06146.

Jia G, Fu Y, Zhao X, Dai Q, Zheng G, Yang Y, et al. N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat Chem Biol. 2011;7(12):885–7. doi:10.1038/nchembio.687.

Jia G, Fu Y, He C. Reversible RNA adenosine methylation in biological regulation. Trends Genet. 2013;29(2):108–15. doi:10.1016/j.tig.2012.11.003.

Kellinger MW, Song CX, Chong J, Lu XY, He C, Wang D. 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. Nat Struct Mol Biol. 2012;19(8):831–3. doi:10.1038/nsmb.2346.

Klimasauskas S, Kumar S, Roberts RJ, Cheng X. HhaI methyltransferase flips its target base out of the DNA helix. Cell. 1994;76(2):357–69.

Klose RJ, Bird AP. Genomic DNA methylation: the mark and its mediators. Trends Biochem Sci. 2006;31(2):89–97. doi:10.1016/j.tibs.2005.12.008.

Kornberg SR, Zimmerman SB, Kornberg A. Glucosylation of deoxyribonucleic acid by enzymes from bacteriophage-infected Escherichia coli. J Biol Chem. 1961;236:1487–93.

Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science. 2009;324(5929):929–30. doi:10.1126/science.1169786.

Kumar S, Cheng X, Klimasauskas S, Mi S, Posfai J, Roberts RJ, et al. The DNA (cytosine-5) methyltransferases. Nucleic Acids Res. 1994;22(1):1–10.

Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet. 2010;11(3):204–20. doi:10.1038/nrg2719.

Lehman IR, Pratt EA. On the structure of the glucosylated hydroxymethylcytosine nucleotides of coliphages T2, T4, and T6. J Biol Chem. 1960;235:3254–9.

Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell. 1992;69(6):915–26.

Lindroth AM, Cao X, Jackson JP, Zilberman D, McCallum CM, Henikoff S, et al. Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. Science. 2001;292(5524):2077–80. doi:10.1126/science.1059745.

Liu Y, Toh H, Sasaki H, Zhang X, Cheng X. An atomic model of Zfp57 recognition of CpG methylation within a specific DNA sequence. Genes Dev. 2012;26(21):2374–9. doi:10.1101/gad.202200.112.

Liu Y, Zhang X, Blumenthal RM, Cheng X. A common mode of recognition for methylated CpG. Trends Biochem Sci. 2013;38(4):177–83. doi:10.1016/j.tibs.2012.12.005.

Liu Y, Olanrewaju YO, Zheng Y, Hashimoto H, Blumenthal RM, Zhang X, et al. Structural basis for Klf4 recognition of methylated DNA. Nucleic Acids Res. 2014;42(8):4859–67. doi:10.1093/nar/gku134.

Low DA, Weyand NJ, Mahan MJ. Roles of DNA adenine methylation in regulating bacterial gene expression and virulence. Infect Immun. 2001;69(12):7197–204. doi:10.1128/IAI.69.12.7197-7204.2001.

Maiti A, Drohat AC. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. J Biol Chem. 2011;286(41):35334–8. doi:10.1074/jbc.C111.284620.

McDonough MA, Loenarz C, Chowdhury R, Clifton IJ, Schofield CJ. Structural studies on human 2-oxoglutarate dependent oxygenases. Curr Opin Struct Biol. 2010;20(6):659–72. doi:10.1016/j.sbi.2010.08.006.

Mierzejewska K, Siwek W, Czapinska H, Kaus-Drobek M, Radlinska M, Skowronek K, et al. Structural basis of the methylation specificity of R.DpnI. Nucleic Acids Res. 2014;42(13):8745–54. doi:10.1093/nar/gku546.

Miyazono K, Furuta Y, Watanabe-Matsui M, Miyakawa T, Ito T, Kobayashi I, et al. A sequence-specific DNA glycosylase mediates restriction-modification in Pyrococcus abyssi. Nat Commun. 2014;5:3178. doi:10.1038/ncomms4178.

Mok YG, Uzawa R, Lee J, Weiner GM, Eichman BF, Fischer RL, et al. Domain structure of the DEMETER 5-methylcytosine DNA glycosylase. Proc Natl Acad Sci U S A. 2010;107(45):19225–30. doi:10.1073/pnas.1014348107.

Morales-Ruiz T, Ortega-Galisteo AP, Ponferrada-Marin MI, Martinez-Macias MI, Ariza RR, Roldan-Arjona T. DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases. Proc Natl Acad Sci U S A. 2006;103(18):6853–8. doi:10.1073/pnas.0601109103.

Mullins EA, Shi R, Parsons ZD, Yuen PK, David SS, Igarashi Y, et al. The DNA glycosylase AlkD uses a non-base-flipping mechanism to excise bulky lesions. Nature. 2015;527(7577):254–8. doi:10.1038/nature15728.

Niu Y, Zhao X, Wu YS, Li MM, Wang XJ, Yang YG. N6-methyl-adenosine (m6A) in RNA: an old modification with a novel epigenetic function. Genomics Proteomics Bioinformatics. 2013;11(1):8–17. doi:10.1016/j.gpb.2012.12.002.

Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature. 2007;448(7154):714–7. doi:10.1038/nature05987.

Ortega-Galisteo AP, Morales-Ruiz T, Ariza RR, Roldan-Arjona T. Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks. Plant Mol Biol. 2008;67(6):671–81. doi:10.1007/s11103-008-9346-0.

Pais JE, Dai N, Tamanaha E, Vaisvila R, Fomenkov AI, Bitinaite J, et al. Biochemical characterization of a Naegleria TET-like oxygenase and its application in single molecule sequencing of 5-methylcytosine. Proc Natl Acad Sci U S A. 2015;112(14):4316–21. doi:10.1073/pnas.1417939112.

Pfaffeneder T, Spada F, Wagner M, Brandmayr C, Laube SK, Eisen D, et al. Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. Nat Chem Biol. 2014;10(7):574–81. doi:10.1038/nchembio.1532.

Ponferrada-Marin MI, Roldan-Arjona T, Ariza RR. ROS1 5-methylcytosine DNA glycosylase is a slow-turnover catalyst that initiates DNA demethylation in a distributive fashion. Nucleic Acids Res. 2009;37(13):4264–74. doi:10.1093/nar/gkp390.

Ponferrada-Marin MI, Martinez-Macias MI, Morales-Ruiz T, Roldan-Arjona T, Ariza RR. Methylation-independent DNA binding modulates specificity of Repressor of Silencing 1 (ROS1) and facilitates demethylation in long substrates. J Biol Chem. 2010;285(30):23032–9. doi:10.1074/jbc.M110.124578.

Ponferrada-Marin MI, Parrilla-Doblas JT, Roldan-Arjona T, Ariza RR. A discontinuous DNA glycosylase domain in a family of enzymes that excise 5-methylcytosine. Nucleic Acids Res. 2011;39(4):1473–84. doi:10.1093/nar/gkq982.

Qian W, Miki D, Zhang H, Liu Y, Zhang X, Tang K, et al. A histone acetyltransferase regulates active DNA demethylation in Arabidopsis. Science. 2012;336(6087):1445–8. doi:10.1126/science.1219416.

Rajakumara E, Law JA, Simanshu DK, Voigt P, Johnson LM, Reinberg D, et al. A dual flip-out mechanism for 5mC recognition by the Arabidopsis SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation in vivo. Genes Dev. 2011;25(2):137–52. doi:10.1101/gad.1980311.

Ramsahoye BH, Biniszkiewicz D, Lyko F, Clark V, Bird AP, Jaenisch R. Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc Natl Acad Sci U S A. 2000;97(10):5237–42.

Roberts RJ, Myers PA, Morrison A, Murray K. A specific endonuclease from Haemophilus haemolyticus. J Mol Biol. 1976;103(1):199–208.

Roberts RJ, Vincze T, Posfai J, Macelis D. REBASE–a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res. 2015;43(Database issue):D298–9. doi:10.1093/nar/gku1046.

Robertson KD. DNA methylation and human disease. Nat Rev Genet. 2005;6(8):597–610. doi:10.1038/nrg1655.

Song J, Rechkoblit O, Bestor TH, Patel DJ. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. Science. 2011;331(6020):1036–40. doi:10.1126/science.1195380.

Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PW, Bauer C, et al. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. Cell. 2013;152(5):1146–59. doi:10.1016/j.cell.2013.02.004.

Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, et al. Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nat Struct Mol Biol. 2014;21(1):64–72. doi:10.1038/nsmb.2735.

Suetake I, Miyazaki J, Murakami C, Takeshima H, Tajima S. Distinct enzymatic properties of recombinant mouse DNA methyltransferases Dnmt3a and Dnmt3b. J Biochem. 2003;133(6):737–44.

Sukackaite R, Grazulis S, Tamulaitis G, Siksnys V. The recognition domain of the methyl-specific endonuclease McrBC flips out 5-methylcytosine. Nucleic Acids Res. 2012;40(15):7552–62. doi:10.1093/nar/gks332.

Sutherland E, Coe L, Raleigh EA. McrBC: a multisubunit GTP-dependent restriction endonuclease. J Mol Biol. 1992;225(2):327–48.

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2009;324(5929):930–5. doi:10.1126/science.1170116.

Tate PH, Bird AP. Effects of DNA methylation on DNA-binding proteins and gene expression. Curr Opin Genet Dev. 1993;3(2):226–31.

Vairapandi M, Duker NJ. Enzymic removal of 5-methylcytosine from DNA by a human DNA-glycosylase. Nucleic Acids Res. 1993;21(23):5323–7.

Vairapandi M, Duker NJ. Partial purification and characterization of human 5-methylcytosine-DNA glycosylase. Oncogene. 1996;13(5):933–8.

Vairapandi M, Liebermann DA, Hoffman B, Duker NJ. Human DNA-demethylating activity: a glycosylase associated with RNA and PCNA. J Cell Biochem. 2000;79(2):249–60.

Wang H, Guan S, Quimby A, Cohen-Karni D, Pradhan S, Wilson G, et al. Comparative characterization of the PvuRts1I family of restriction enzymes and their application in mapping genomic 5-hydroxymethylcytosine. Nucleic Acids Res. 2011;39(21):9294–305. doi:10.1093/nar/gkr607.

Wang L, Zhou Y, Xu L, Xiao R, Lu X, Chen L, et al. Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex. Nature. 2015;523(7562):621–5. doi:10.1038/nature14482.

Warren RA. Modified bases in bacteriophage DNAs. Annu Rev Microbiol. 1980;34:137–58. doi:10.1146/annurev.mi.34.100180.001033.

Wilson GG, Murray NE. Restriction and modification systems. Annu Rev Genet. 1991;25:585–627. doi:10.1146/annurev.ge.25.120191.003101.

Woo HR, Pontes O, Pikaard CS, Richards EJ. VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. Genes Dev. 2007;21(3):267–77. doi:10.1101/gad.1512007.

Wu JC, Santi DV. On the mechanism and inhibition of DNA cytosine methyltransferases. Prog Clin Biol Res. 1985;198:119–29.

Wu JC, Santi DV. Kinetic and catalytic mechanism of HhaI methyltransferase. J Biol Chem. 1987;262(10):4778–86.

Wyatt GR, Cohen SS. The bases of the nucleic acids of some bacterial and animal viruses: the occurrence of 5-hydroxymethylcytosine. Biochem J. 1953;55(5):774–82.

Yoder JA, Soman NS, Verdine GL, Bestor TH. DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe. J Mol Biol. 1997;270(3):385–95. doi:10.1006/jmbi.1997.1125.

Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, et al. The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell. 2013;153(1):193–205. doi:10.1016/j.cell.2013.02.033.

Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, et al. Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis. Cell. 2006;126(6):1189–201. doi:10.1016/j.cell.2006.08.003.

Zhang L, Lu X, Lu J, Liang H, Dai Q, Xu GL, et al. Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. Nat Chem Biol. 2012;8(4):328–30. doi:10.1038/nchembio.914.

Zhang G, Huang H, Liu D, Cheng Y, Liu X, Zhang W, et al. N6-methyladenine DNA modification in Drosophila. Cell. 2015;161(4):893–906. doi:10.1016/j.cell.2015.04.018.

Zheng X, Pontes O, Zhu J, Miki D, Zhang F, Li WX, et al. ROS3 is an RNA-binding protein required for DNA demethylation in Arabidopsis. Nature. 2008;455(7217):1259–62. doi:10.1038/nature07305.

Zheng G, Dahl JA, Niu Y, Fedorcsak P, Huang CM, Li CJ, et al. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol Cell. 2013;49(1):18–29. doi:10.1016/j.molcel.2012.10.015.

Zhong X, Du J, Hale CJ, Gallego-Bartolome J, Feng S, Vashisht AA, et al. Molecular mechanism of action of plant DRM de novo DNA methyltransferases. Cell. 2014;157(5):1050–60. doi:10.1016/j.cell.2014.03.056.

Zhou T, Xiong J, Wang M, Yang N, Wong J, Zhu B, et al. Structural basis for hydroxymethylcytosine recognition by the SRA domain of UHRF2. Mol Cell. 2014;54(5):879–86. doi:10.1016/j.molcel.2014.04.003.
Zhu JK. Active DNA, demethylation mediated by DNA glycosylases. Annu Rev Genet. 2009;43:143–66. doi:10.1146/annurev-genet-102108-134205.
Zhu C, Yi C. Switching demethylation activities between AlkB family RNA/DNA demethylases through exchange of active-site residues. Angew Chem Int Ed Engl. 2014;53(14):3659–62. doi:10.1002/anie.201310050.

Current and Emerging Technologies for the Analysis of the Genome-Wide and Locus-Specific DNA Methylation Patterns

Jörg Tost

Abstract

DNA methylation is the most studied epigenetic modification, and altered DNA methylation patterns have been identified in cancer and more recently also in many other complex diseases. Furthermore, DNA methylation is influenced by a variety of environmental factors, and the analysis of DNA methylation patterns might allow deciphering previous exposure. Although a large number of techniques to study DNA methylation either genome-wide or at specific loci have been devised, they all are based on a limited number of principles for differentiating the methylation state, viz., methylation-specific/methylation-dependent restriction enzymes, antibodies or methyl-binding proteins, chemical-based enrichment, or bisulfite conversion. Second-generation sequencing has largely replaced microarrays as readout platform and is also becoming more popular for locus-specific DNA methylation analysis. In this chapter, the currently used methods for both genome-wide and locus-specific analysis of 5-methylcytosine and as its oxidative derivatives, such as 5-hydroxymethylcytosine, are reviewed in detail, and the advantages and limitations of each approach are discussed. Furthermore, emerging technologies avoiding PCR amplification and allowing a direct readout of DNA methylation are summarized, together with novel applications, such as the detection of DNA methylation in single cells or in circulating cell-free DNA.

J. Tost
Laboratory for Epigenetics and Environment, Centre National de Génotypage, CEA-Institut de Génomique, 2 rue Gaston Crémieux CP5721, Evry 91000, France
e-mail: tost@cng.fr

© Springer International Publishing Switzerland 2016

A. Jeltsch, R.Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*,
Advances in Experimental Medicine and Biology 945,
DOI 10.1007/978-3-319-43624-1_15

Abbreviations

5caC	5-Carboxylcytosine
5fC	5-Formylcytosine
5hmC	5-Hydroxymethylcytosine
5mC	5-Methylcytosine
Aba-seq	AbaSI-coupled sequencing
Ccf	DNA circulating cell-free DNA
ChIP	Chromatin immunoprecipitation
CMS	5-Cytosinemethylenesulfonate
COBRA	Combined bisulfite restriction analysis
COLD	Coamplification at lower denaturation temperature
ddPCR	Digital droplet PCR
DREAM	Digital restriction enzyme analysis of methylation
FFPE	Formalin fixed paraffin embedded
GLIB	Glucosylation, periodate oxidation, biotinylation
HELP	HpaII tiny fragment enrichment by ligation-mediated PCR
HELP-GT	HpaII tiny fragment enrichment by ligation-mediated PCR-glycosyl *t*ransferase assay
hMeSEAL	5hmC-selective chemical labeling
M	Million
MALDI-TOF-MS/MALDI-MS	Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry
MBD	Methyl-binding domain
MSDK	Methylation-specific digital karyotyping
MeDIP	Methylated DNA immunoprecipitation
MIRA	Methylated-CpG island recovery assay
MRE	Methylation-specific restriction enzyme
MS	Methylation sensitive
MSCC	Methylation-specific cut counting
MS-FLAG	Methylation-sensitive fluorescent amplicon generation
MS-HRM	Methylation-specific high-resolution melting analysis
MS-MLPA	Methylation-specific multiplexed ligation probe amplification
MSP	Methylation-specific PCR
MS-SNuPE	Methylation-specific single-nucleotide primer extension
NGS	Next-/second-generation sequencing
OxBS	Oxidative bisulfite
PBAT	Post-bisulfite adaptor tagging
PBMC	Peripheral blood mononuclear cells
QAMA	Quantitative analysis of methylated alleles
RRBS	Reduced representation bisulfite sequencing
RRHP	Reduced representation 5-hydroxymethylcytosine profiling
SBS	Sequencing by synthesis
SCAN	Single chromatin molecule analysis at the nanoscale

SMART-MSP	Sensitive melting analysis after real-time methylation-specific PCR
SMRT	Single-molecule real time
SuBLiME	Streptavidin bisulfite ligand methylation enrichment
TAB-seq	TET-assisted bisulfite sequencing
TET	Ten-eleven translocation (enzyme)

1 Introduction

Epigenetic phenomena are mediated by a variety of molecular mechanisms including posttranscriptional histone modifications, histone variants, ATP-dependent chromatin remodeling, and small and other noncoding RNAs and DNA methylation (Tost 2008a). These diverse molecular mechanisms are all closely intertwined and stabilize each other to ensure the faithful propagation of an epigenetic state over time and especially through cell division. Epigenetics and the analysis of epigenetic modifications have come to a central stage for many developmental and biomedical questions, and the advent of second-/next-generation sequencing (NGS) has revolutionized the way of interrogating the epigenome.

In this chapter, I will provide a comprehensive overview of the technologies applied for the genome-wide and locus-specific analysis of DNA methylation, the best-studied epigenetic modification. As the description of all methods used for the different applications of DNA methylation analysis is beyond the scope of this chapter, I will concentrate on the most widely used methods, powerful emerging concepts, as well as approaches that are or have the potential to be used in the clinics.

1.1 DNA Methylation

In mammals, DNA methylation is the most prevalent DNA modification and is almost entirely found on the C5 position of the pyrimidine ring of cytosines in the context of CpG dinucleotides (Bird 2002). 5-Methylcytosine accounts for ~1 % of all bases, whereby the majority (75 %) of CpG dinucleotides throughout mammalian genomes is methylated. Methylation of cytosines in the context of CpHpG or CpHpH sequences (where H is A, C, or T) has been detected in embryonic stem cells, the brain, and plants but is rarely found in other somatic mammalian/human tissues. DNA methylation can have profound effects on gene expression, is crucial for proper development, and is implicated in disease processes, particularly tumorigenesis (Baylin and Jones 2011), but more recently it also got connected to other complex diseases, including autoimmune and inflammatory diseases and neurodegenerative, psychiatric, and metabolic disorders (Absher et al. 2013; Miceli-Richard et al. 2016; Karatzas et al. 2014; Chandra et al. 2015; Ronn and Ling 2015; Voelter-Mahlknecht 2016; Yu et al. 2015a). DNA methylation might act as a memory of both internal and external environmental influences to which the cells of an organism have been exposed. The variety of epigenetic modifications in mammalian DNA has recently further increased with the

discovery of 5-hydroxymethylcytosine (5hmC) and its oxidative derivatives, 5-formyl (5fC) and 5-carboxylcytosine (5caC) in mammalian neurons and embryonic stem cells, which are formed from 5-methylcytosines by a catalytic oxidation mediated by the TET proteins (Kriaucionis and Heintz 2009; Tahiliani et al. 2009). The 5hmC modification has been found in nearly all mouse embryonic tissues, but with the exception of brain tissue and bone marrow only very low levels (0.2–0.05 % of all cytosines) are detectable in adult tissue (Ruzov et al. 2011b; Globisch et al. 2010).

As a covalent DNA-based modification, which is technically relatively easy to investigate, DNA methylation has been intensively studied since the 1980s and is the best-studied epigenetic mark. For the analysis of DNA methylation patterns, sensitive and quantitative methods are required to detect even subtle changes in the degree of methylation, as biological samples often represent a heterogeneous mixture of different cells, for example, tumor and non-tumor cells from tissue biopsies or DNA extracted from peripheral blood mononuclear cells (PBMCs). Realizing the importance of epigenetic changes in development and disease, a large number of technologies for the study of DNA methylation have been developed in the last years. However, no single method has emerged as the "gold" standard technique unifying cost-efficiency, throughput, quantitative accuracy and sensitivity, possibility for whole-genome analysis, and precise investigation of individual CpG positions. Therefore, the choice of the method mainly depends on the required answer to a fundamental or biomedical research problem. A major advantage of DNA methylation analysis is that it can be carried out on DNA isolated from nearly all biological tissues or body fluids, as DNA methylation marks seem to be stable even under prolonged storage conditions. It has, for example, been shown that DNA from the blood spots of Guthrie cards can be reliably extracted for up to 20 years with a quality suitable for genome-wide DNA methylation analyses (Ghantous et al. 2014). Archived specimens fixed in a variety of fixatives are also suitable for DNA methylation analysis, but to a different degree. Despite suffering from large damage and modification of the nucleic acids (Srinivasan et al. 2002), DNA from *f*ormalin-*f*ixed *p*araffin-*e*mbedded (FFPE) samples can be used for locus-specific analyses of DNA methylation patterns using, for example, pyrosequencing (Leong et al. 2013), or after a ligation and amplification-based repair step for genome-wide analysis using the Infinium BeadChips (Thirlwell et al. 2010). However, alternative fixatives such as PAXgene-based preservation proved superior in terms of accuracy and reproducibility compared to FFPE specimens (Andersen et al. 2014).

Although DNA methylation is technically easier to handle and the requirements for pre-analytical sample processing are less stringent compared to the analysis of RNA and protein modifications, several challenges are nonetheless associated with the analysis of DNA methylation patterns in tissue. Biological specimens are composed of a large number of cell types, each associated with its own DNA methylome, and appropriate control of the cellular composition needs to be carried out using either careful pathological examination, complementary molecular methods, e.g., for assessing the tumor percentage (Van Loo et al. 2010), or statistical methods, e.g., for blood cell composition (Houseman et al. 2012). These points are of course also valid for the analysis of RNA- or protein-based biomarkers. Heterogeneity at the level of DNA methylation, which is transformed into polymorphisms of the

target sequence, might also be problematic for some of the technologies used for the assessment of DNA methylation levels in a candidate region, when primers fail to amplify, e.g., partially methylated molecules (Mikeska et al. 2010; Alnaes et al. 2015). Another yet unanswered question is under which conditions and to which extent an accessible tissue such as blood, urine, or saliva can be used as a surrogate for the target organ, if the primary disease organ is not available for analysis.

2 Principles of DNA Methylation Detection

Several generations of genome-wide methods for DNA methylation analysis have been developed adapting to different analytical supports with increasing levels of resolution and coverage. Except for the highly popular epigenotyping arrays, NGS-based technologies have largely replaced microarrays as the readout platform for DNA methylation analysis, and the latter are therefore not covered in this review. Historical methods for DNA methylation analysis have been reviewed in Tost (2008b and Harrison and Parle-McDermott (2011).

Despite the advancement of readout platforms, the main approaches for the discrimination of methylation have so far little evolved, although new methods allowing potentially a direct readout of DNA methylation patterns have been devised (Song et al. 2012; Flusberg et al. 2010; Clarke et al. 2009). Current assays are based on four main principles:

1. The use of methylation-specific restriction endonucleases, i.e., enzymes that are blocked by methylated cytosines in their recognition sequence (Bird and Southern 1978), which are widely used for the analysis of methylation patterns in combination with their methylation-insensitive isoschizomers. Although methods based on methylation-specific restriction enzymes are simple and cost-effective and might be able to distinguish between methylcytosine and its oxidative derivatives, they are hampered by the limitation to specific restriction sites, as only CpG sites found within the recognition sequences of restriction enzymes can be analyzed (Fazzari and Greally 2004). Information complementary to that obtained by methylation-specific restriction digests can be obtained by methylation-dependent restriction enzymes such as McrBC, which cleaves between two non-palindromic G/A^{me}C sites 40–3000 bp apart from each other with an optimal distance of 55–103 bp (Stewart et al. 2000).
2. The methylated fraction of a genome can be enriched by precipitation with a bead-immobilized antibody specific for 5-methylcytosine following a protocol similar to chromatin immunoprecipitation (ChIP) and analyzed on microarrays (Weber et al. 2005) or by sequencing (Down et al. 2008; Feber et al. 2011).
3. Similar results can be obtained by affinity purification of methylated DNA with methyl-binding domain (MBD) proteins such as MeCP2 (Brinkman et al. 2010) or MBD2 in combination with MBD3L1 (Rauch and Pfeifer 2005).
4. The most widely used approach consists of the chemical modification of genomic DNA with sodium bisulfite. This chemical reaction induces hydrolytic deamination of unmethylated cytosines to uracils, while methylated cytosines are resistant

to the conversion under the chosen reaction conditions (Frommer et al. 1992; Shapiro et al. 1974). This method translates the methylation signal into a sequence difference. After performing PCR, the methylation status at a given position is manifested in the ratio of C (former methylated cytosine) to T (former unmethylated cytosine) and can be analyzed as a virtual C/T polymorphism spanning the entire allele frequency spectrum from 0 to 100 % in the bisulfite-treated DNA. A number of commercial kits have been developed, which allow the user-friendly conversion of genomic DNA from various sources. Nonetheless, the chemical treatment degrades a significant amount of the input DNA leading to the loss of a substantial amount of the starting material that could become problematic if only a very limited amount is available (Holmes et al. 2014). Specific quality control assays have been developed, which allow for the accurate quantification of total, amplifiable converted and unconverted DNA after bisulfite treatment (Campan et al. 2009). Furthermore, the difference in GC content for different molecules induced by the bisulfite conversion and their former DNA methylation status might influence their amplification efficiency and optimal annealing temperature. Therefore, the preferable amplification of unmethylated or methylated DNA is a common complication for methods including PCR amplification of bisulfite-treated DNA (Warnecke et al. 1997; Grunau et al. 2001). Several methods have been devised for the potential correction of the amplification bias, including single-molecule PCR (Chhibber and Schroeder 2008), addition of betaine to the PCR (Voss et al. 1998), inclusion of CpGs in the sequence of the amplification primers (Wojdacz and Hansen 2006), and/or variation of the annealing temperature (Shen et al. 2007). Nonetheless, all these methods cannot be generally applied but have to be carefully tested using standards with a known degree of DNA methylation at the specific locus of interest. Another important problem might arise from clonal amplification during the PCR following bisulfite conversion, leading to an ostensible faultless readout, which is however not representing the true distribution of DNA methylation in the original sample (Zhang and Jeltsch 2010). Technical replicates of the bisulfite conversion or at least of the PCR amplification or the addition of molecule-specific barcodes during the first-strand synthesis might help to detect such problems (Miner et al. 2004; Zhang and Jeltsch 2010). It should also be noted that standard bisulfite conversion protocols cannot discriminate between 5-hydoxymethylcytosine and 5-methylcytosine, which are converted with a similar efficiency. Specialized protocols such as oxidative bisulfite sequencing (Booth et al. 2012) or TAB-seq (Yu et al. 2012) are required to avoid any potential influence of concomitant hydroxymethylation on accurate quantification of DNA methylation. Methods for the specific analysis of 5-hydroxymethylation are detailed in Sect. 15.10 of this chapter.

Depending on the requirement for resolution, coverage, quantification, and throughput, these four main assay principles have been combined with PCR-, microarray-, or sequencing-based readout technologies. Nowadays, NGS methods have largely replaced microarray-based readouts, because they allow a truly genome-wide discovery of differentially methylated CpGs and/or regions, they require in

most cases lower starting amounts of DNA, and they limit DNA amplification during the analysis, thereby reducing potential amplification biases.

3 Principles of Next-Generation Sequencing

Illumina's sequencing-by-synthesis (SBS) chemistry is currently the most widely adopted chemistry in the field. The sequencing library is prepared by random fragmentation of the DNA or cDNA sample, followed by 5′ and 3′ adapter ligation. Multiple samples can be analyzed in parallel using barcodes, which are short six to eight base pair sequences specific for a given sample that can be incorporated into one or both (dual indexing) adapters. Alternatively, the tagmentation method combines the fragmentation and ligation reactions into a single step. For cluster generation, the library is loaded into a flow cell where fragments are captured on surface-bound oligos complementary to the library adapters. Each fragment is then amplified into a distinct, clonal cluster through bridge amplification. Sequencing utilizes four reversible terminator-bound dNTPs, which are present during each sequencing cycle that - like in Sanger sequencing – detect single bases as they are incorporated into DNA template strands. Sequencing can be carried out in single or paired-end modus, the latter giving more precise location about the sequence fragment analyzed. A variety of sequencing machines with different throughput capacities are available from Illumina, ranging from the Mini-seq system (output 2–7 Gb) to the HiSeq X Ten platform (output 16–18 Tb), allowing to sequence between 50 and 600 base pairs per fragment. Recent improvements include the use of ordered flow cells optimizing the number of detectable clusters and improved detection systems reducing the run time threefold. Further, the HiSeq X Five and X Ten allow for the first time a cost-efficient whole-genome bisulfite sequencing of mammalian genomes (see below), while the MiSeq system has received regulatory approval for diagnostic applications, allowing potentially to transfer the targeted analysis of DNA methylation to a diagnostic platform.

For sequencing on the SOLiD™ System (ThermoFisher), (c)DNA is fragmented and clonal bead populations are prepared in microreactors containing template, PCR reaction components, beads, and primers. After PCR, beads with extended templates are enriched, and the template on the selected beads undergoes a 3′ modification to allow covalent attachment to the flow cell. In contrast to the Illumina protocol, a set of four fluorescently labeled di-base probes compete for ligation to the sequencing primer. Specificity of the di-base probe is achieved by interrogating every first and second base in each ligation reaction. Multiple cycles of ligation, detection, and cleavage are performed with the number of cycles determining the eventual read length. Following a series of ligation cycles, the extension product is removed, and the template is reset with a primer complementary to the n-1 position for a second round of ligation cycles. As most bases are interrogated in two independent ligation reactions by two different primers, the accuracy of the sequencing is very high. However, the output is much lower than on Illumina sequencers and the price per sequenced base substantially higher, which has limited its use. In the field of DNA methylation analysis, SOLiD systems have mainly been used for the sequencing of affinity-enriched libraries.

The Ion Torrent sequencing platform (ThermoFisher) is a semiconductor-based sequencing exploiting the release of hydrogen ions upon nucleotide incorporation during the sequencing-by-synthesis reaction (Rothberg et al. 2011). This approach has some similarities to the pyrosequencing reaction monitoring the release of pyrophosphate, which was used in the Roche/454 sequencing systems, which are currently phased out. While the Ion Torrent and GS Junior benchtop sequencers require less upfront investment and have shorter run times compared to some earlier Illumina sequencers, they do not allow for the same throughput and their error rate might be slightly higher.

If not explicitly otherwise stated, all sequencing approaches described in this chapter have been performed on the different Illumina platforms (GAIIx, MiSeq, HiSeq), and while the general principles of each method are transferrable to any sequencing platform, protocols might need to be adapted to satisfy platform-specific requirements.

4 Global Methylation Content of a Sample

Methods for the analysis of global DNA methylation levels in a sample determine the overall 5-methylcytosine content or its changes affecting the entire epigenome, respectively. They do, however, not give any information about the location or repartition of DNA methylation on the genome. 5-Methylcytosine can be differentiated from its unmethylated counterpart by the different mass or polarity of the two bases (cytosine and 5-methylcytosine), which can be used for chromatographic or mass spectrometric separation (Wiebers 1976; Eick et al. 1983; Fraga et al. 2002). The 5-methylcytosine content is measured after hydrolysis to mononucleosides, and comparison to an internal standard enables quantification. Particularly, since the discovery of the oxidative derivatives of 5-methylcytosine, the mass spectrometry detection has become very popular, which is also due to its exquisite sensitivity and the low amounts of required starting material (Globisch et al. 2010; Le et al. 2011). Less accurate quantification of the global methylation content can also be made with a variety of ELISA-based commercial kits, but these kits are normally much less sensitive than mass spectrometric or chromatographic methods enabling only the detection of very large changes in the DNA methylation content.

Bacterial methyltransferases, e.g., M.SssI, transfer a methyl group from the universal methyl donor *S*-adenosyl-L-methionine to unmethylated CpG positions. The methyl acceptor assay makes use of these enzymes and analyzes the amount of incorporated radiolabeled methyl groups into a sample (Bestor and Ingram 1983). The measured amount of radioactive label correlates thus inversely with the degree of its methylation prior to labeling. Similarly, the cytosine extension assay combines methylation-specific restriction digestion and single-nucleotide extension with radio- or fluorescently labeled dCTP complementary to the guanine 5′ overhang created by the digestion (Pogribny et al. 1999; Bönsch et al. 2004). The pyrosequencing-based Luminometric Methylation Assay (LUMA) is based on the differential digestion of a sample with a methylation-specific endonuclease or

its methylation-insensitive isoschizomer and the successive dispensation of four nucleotides complementary to the overhang created by the endonucleases (Karimi et al. 2006). The pyrosequencing-based analysis of repetitive elements such as Alu or LINE1 has also been widely used as surrogate for the global DNA methylation level (Yang et al. 2004). Methods analyzing the total amount of 5-methylcytosine in a sample are used to analyze and follow global DNA methylation changes induced by demethylating pharmaceutical agents in patients with hematological malignancies at various time points of treatment (Mund et al. 2005; Liu et al. 2007; Kantarjian et al. 2006), to investigate the efficacy of novel demethylating agents (Balch et al. 2005), to detect and predict the outcome of various human cancers (Hur et al. 2014; Inamura et al. 2014), as well as to detect and quantify the effect of environmental exposure on the DNA methylome (Bollati et al. 2007; Marques-Rocha et al. 2016).

In situ hybridization methods with antibodies directed against 5-methylcytosine or its oxidative derivatives allow the measurement of the methylation content and its potentially cell-type-specific distribution. They can be used to visualize the modified bases in the cell nuclear context (Miller et al. 1974; Rougier et al. 1998; Salvaing et al. 2015). As only clustered methylated CpGs (e.g., in repeat elements) can be recognized at the chromosomal level, methylation patterns at relatively small loci, such as CpG islands, contribute little to the overall staining profile. Therefore, technologies using an ultra-sensitive and rapid fluorescence scanning system with a sub-micrometer resolution have been devised to achieve the detection of methyl groups at specific promoters isolated from genomic DNA by restriction digestion and hybridization to capture oligonucleotides immobilized on a glass slide (Pröll et al. 2006).

5 Whole Methylome Analyses

Although potentially confounded by the presence of 5-hydroxymethylcytosine, whole-genome bisulfite sequencing or MethylC-seq can be considered as the current gold standard for the genome-wide identification of differentially methylated regions at single-nucleotide resolution. This technology is currently used in a number of international large-scale projects to map the methylome of various human tissues and cell types (Adams et al. 2012; Roadmap Epigenomics et al. 2015; Schultz et al. 2015). It overcomes the limitations of cloning and Sanger sequencing, a low-throughput method limited to a small number of loci of interest, in which the quantitative resolution was limited by the number of clones analyzed (in most studies <20). Furthermore, whole-genome bisulfite sequencing avoids problems with the primer design that often introduces multiple biases in locus-specific studies (Grunau et al. 2001; Warnecke et al. 2002). However, the unprecedented quantitative and spatial resolution that is currently transforming DNA methylation analysis comes at a high cost, as it requires substantial sequencing to obtain a proper and even coverage and necessitates bioinformatic expertise and resources. Although low-coverage bisulfite sequencing can yield some information about global DNA methylation alterations, it does not yield reliable locus-specific information (Popp et al. 2010).

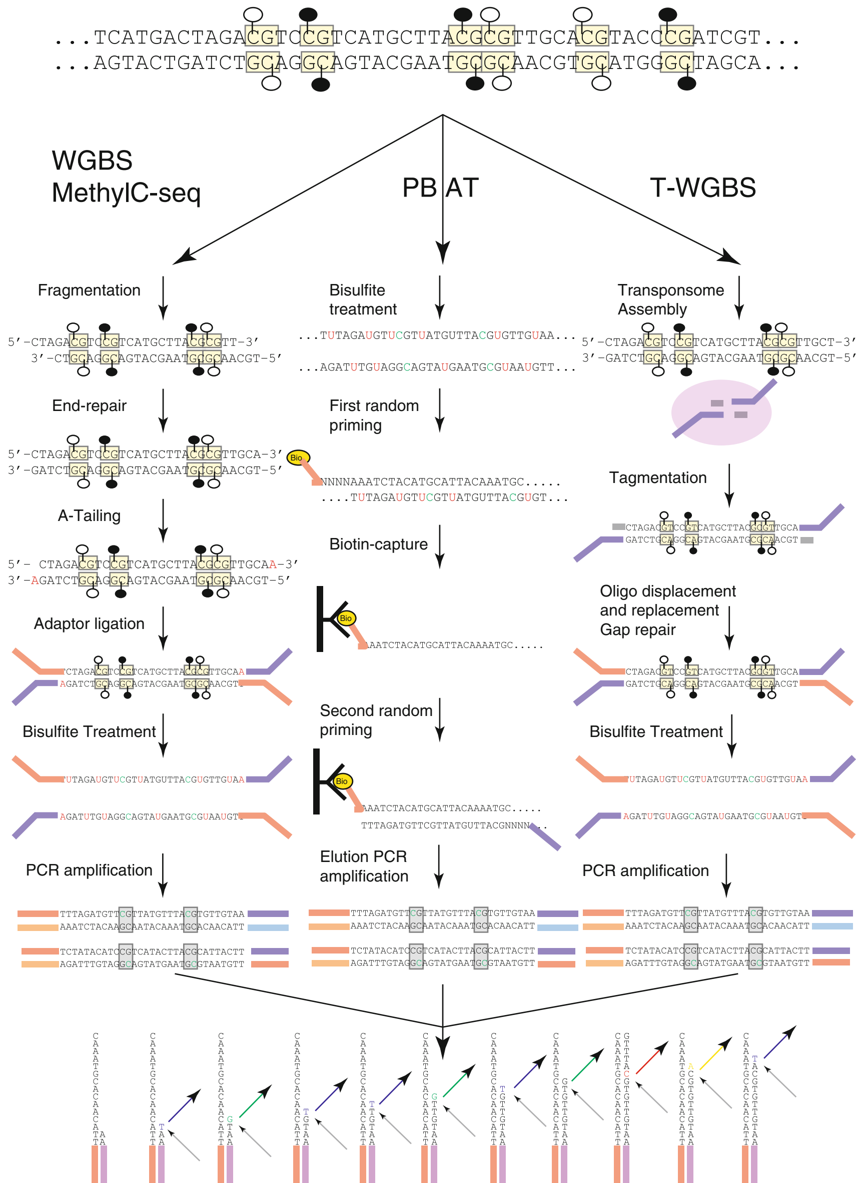


Fig. 1 Schematic outline of the three most commonly used approaches for whole-genome bisulfite sequencing (from *left to right*): MethylC-seq, PBAT, and tagmentation-based WGBS. Details of the procedures are described in the text

The most widely used protocol (Fig. 1) consists of the fragmentation of genomic DNA, adapter ligation, bisulfite conversion, and limited amplification using adapter-specific PCR primers. While initially several micrograms of DNA were required to perform whole-genome bisulfite sequencing, the replacement of electrophoresis and

gel extraction steps by magnetic beads has enabled generation of libraries suitable for sequencing from ~ 100 ng of input material (Urich et al. 2015). Libraries have been reported to be constructed from even less input material but require in most cases a high number of PCR cycles (up to 25 cycles (Kobayashi et al. 2012)), inducing potentially a large bias for the estimation of the DNA methylation levels, and require a much more substantial sequencing effort to obtain a homogeneous and sufficient coverage. To assess the bisulfite conversion efficiency, DNA of the bacteriophage λ, only containing unmethylated cytosines, is spiked in the reaction. Mapping the reads against the bisulfite-converted genome of the phage and counting any remaining cytosines allow then to identify problems of bisulfite conversion and estimate the conversion rates. A large number of programs have been developed to perform the quality control, preprocessing steps (such as adaptor, barcode, and quality score trimming), mapping the reads to a bisulfite-converted reference genome, scoring of DNA methylation levels (count statistics), and identification of differentially methylated CpG positions (DMCs) and regions (DMRs) (Adusumalli et al. 2015). The protocol has been widely used for the methylome analysis of a large number of organisms and plants, as well as human tissues, and ~90–95 % of the cytosines present in the genome are routinely covered (Lister et al. 2008, 2009, 2011; Li et al. 2010b; Chalhoub et al. 2014; Lyko et al. 2010; Guo et al. 2014).

While most commonly performed on the Illumina sequencing platform, which allows for a much higher coverage, protocols and analytical pipelines have also been devised for the SOLiD platform (Kreck et al. 2012; Bormann Chung et al. 2010), and they have been applied in some studies (Kreck et al. 2013; Hansen et al. 2011).

As the Watson and Crick strands of the DNA are no longer complementary after bisulfite conversion, after PCR amplification and synthesis of the DNA complementary to either the Watson or the Crick strand, a bisulfite-converted genome contains four strands. While methylation is mostly symmetrical in the input DNA, the MethylC-seq protocol will analyze only one of the two strands. A variation of the MethylC-seq protocol, which allows capturing all four strands of a bisulfite-treated genomic DNA using an alternative sequencing adaptor strategy, has been devised (BS-seq; Cokus et al. 2008). While allowing a more comprehensive mapping of cytosines, this strategy requires a large amount of input DNA and comes at the cost of a more complex bioinformatic analysis. Therefore, this approach has been rarely used (Popp et al. 2010). Similarly, whole-genome pre-amplification of bisulfite-treated DNA has been proposed as an alternative to obtain sufficient material for sequencing (Kobayashi et al. 2012); however, the accuracy and reproducibility of the whole-genome amplification are still under debate, as it is prone to bias, especially if low amounts of input DNA are used (Bundo et al. 2012).

Tagmentation is based on a hyperactive variant of the prokaryotic Tn5 transposase that randomly fragments DNA and tags ends with the sequencing adaptors that can be subsequently used for amplification. While initially used for low-input genome sequencing (Adey et al. 2010), this approach has been adapted to the analysis of genome-wide DNA methylation analysis (Tn5mC-seq or T-WGBS (Adey and Shendure 2012; Wang et al. 2013b)) (Fig. 1) and single-cell RNA sequencing (Brouilette et al. 2012) and more recently also for the analysis of chromatin

accessibility (ATAC-seq; Buenrostro et al. 2013), transcription factor binding sites, and histone modifications (Schmidl et al. 2015). As the tagmentation requires double-stranded DNA as a target for the transposition reaction, it is performed prior to the bisulfite conversion. Transposase complexes are loaded with methylated oligonucleotide (except for the 19 bp transposase recognition sequence) to retain sequence identify after the bisulfite treatment and enable the use of the standard amplification primers. A second complementary methylated adaptor is added replacing the transposase recognition sequence and ligated using gap repair. Double-stranded DNA fragments are bisulfite converted and subsequently PCR amplified to append the outer flow cell compatible primers. The protocol can be carried out with very little input (down to 10 ng), as the tagmentation step removes the need for the multiple steps of the conventional protocol for library preparation (DNA shearing, 3′end repair, adenylation, and adapter ligation). Furthermore, the presence of unmethylated nucleotides during the gap repair step serves as an internal control for bisulfite conversion efficiency and abolishes the need for DNA spike-ins. The tagmentation reaction has been shown to be little affected by the GC content (Adey and Shendure 2012). Up to 96 % of CpGs can be covered with this approach, and ~70 % of reads align to the genome, a number slightly lower compared to the standard MethylC-seq protocol. However, the coverage has been shown to be slightly more uniform compared to the standard MethylC-seq (Adey and Shendure 2012). Overall, MethylC-seq and T-WGBS do yield highly similar results in terms of methylation levels and coverage of the genome, and there seems to be no sequence bias for the insertion of the transposase (Wang et al. 2013b). Of note, T-WGBS has been found unsuitable for the analysis of DNA extracted from FFPE tissue (Wang et al. 2013b).

A potential drawback of both the MethylC-seq and the tagmentation-based protocols is that the adaptors are ligated to the DNA fragments prior to the bisulfite conversion. The treatment of DNA with sodium bisulfite leads to a substantial degradation of DNA and reduces significantly the amount of amplifiable DNA through the induction of double-stranded breaks between the adaptors. Therefore, protocols performing the adaptor tagging after bisulfite treatment (PBAT: post-bisulfite adaptor tagging) have been devised and shown to enable efficient library construction from as little as 125 pg of DNA (Miura et al. 2012) and even single cells (Smallwood et al. 2014). This approach has therefore been used to sequence the methylomes of rare cell populations, such as primordial germ cells or zygotes (Peat et al. 2014; Kobayashi et al. 2013; Kobayashi et al. 2012; Shirane et al. 2013). Adaptor tagging is performed with two rounds of random primer extension using oligonucleotides with a random tetramer sequences at the 3′-end of the amplification primers containing the Illumina adaptor sequences. Primers are biotinylated to allow capture of the biotinylated fragments after the first-strand synthesis on magnetic beads. When starting from ~100 ng of DNA, this method allows for the routine PCR-free construction of libraries for methylome-wide sequencing (Miura et al. 2012), thereby avoiding the problem of high PCR duplicate rates that is frequently occurring in PBAT protocols due to the preferential binding of the random amplification primers when the library has insufficient complexity and diversity. Furthermore, random primers tend to preferentially amplify sequences with an elevated GC content.

While the classic MethylC-seq tends to cover GC-poor regions better than GC-rich regions, this phenomenon seems to be reversed in the PBAT protocol, suggesting that – if PCR amplification is required – the combination of the two approaches will probably yield the most even coverage of the methylome.

6 Genome-Wide Methylation Analyses Using NGS

Whole methylome analyses covering each CpG in the genome at single-base resolution remain complex and resource intensive when aiming for a reasonable coverage of at least 20–30 × on a given CpG site and are, therefore, not yet feasible in large cohorts. Furthermore, it has been shown that more than half of all reads do not contain even a single CpG dinucleotide and are thus without any information for DNA methylation analysis, making suboptimal use of the sequencing capacities of current sequencers (Ziller et al. 2013). Furthermore, many CpGs will not show variable DNA methylation under any condition. Therefore, several approaches have been developed to concentrate on the "potentially informative" fraction of the genome (Table 1). These approaches make use of either sequence features such as the CpG density or use antibody, protein, or chemical labeling-based methods to enrich the methylated or unmethylated fraction of the genome. However, it should be kept in mind that all of them will only analyze part of the CpGs present in a genome and none does provide a comprehensive analysis of the methylome.

6.1 Bisulfite-Based Methods

*R*educed *r*epresentation *b*isulfite *s*equencing (RRBS; Meissner et al. 2005, 2008; Gu et al. 2011) is currently the most popular alternative to WGBS, as it requires significantly less sequencing, and CpG-rich regions that are enriched by restriction enzyme digestion are relatively well covered (Meissner et al. 2008; Bock et al. 2010; Harris et al. 2010). RRBS makes use of a methylation-insensitive restriction endonuclease with a CG-rich recognition sequence, such as MspI, which cuts between the two Cs in the target sequence CCGG, which is frequently found in CpG islands and promoter regions. After a size selection step, the generated DNA fragments are subsequently used for a standard library construction using methylated adaptors followed by bisulfite conversion. RRBS interrogates approximately 80 % of CpG islands and 60 % of promoters; this corresponds to only ~12 % of the 28 million (M) CpGs in a human genome, mainly in regions of high CpG density, such as CpG islands, and cannot be designed to target specific genomic regions. On the other hand, RRBS can be performed from minute amounts of DNA, and especially for nonhuman samples, for which no Infinium Methylation BeadChips exist, RRBS is probably the most cost-efficient method to obtain comprehensive high-resolution and quantitative methylome data, with as little as 20 M sequencing reads approaching saturation (Bock et al. 2010). However, nonuniform coverage of CpGs across samples might be an issue, and many CpG poor regions might be missed.

Table 1 Key characteristics of most widely used methods for the genome-wide analysis of DNA methylation

Technology	The technology behind	*M*ethylated/ *u*nmethylated fraction	Single-nucleotide resolution (yes/no)	DNA input in µg	Required number of reads per sample in M	Coverage of the human genome	Coverage of CpGs in the human genome	Comments	References
WGBS/ MethylC-seq	Adaptor ligation before or after bisulfite treatment of genomic DNA prior to sequencing	U + M	Yes	0.1–5	500	95 %	29 M	Comprehensive coverage of all cytosines in a genome, several variations of library preparation exist, cost-intensive, but will become the gold standard for the discovery technologies with decreasing sequencing costs, no distinction between 5mC and 5hmC	Urich et al. (2015), Urich et al. (2013b), Miura et al. (2012)
RRBS	Restriction digest allows for size selection of small fragments, used for library construction, bisulfite conversion, and sequencing	U + M	Yes	0.01–1	20	4 %	1.6 M	85 % of CpG islands are covered, cost-effective sequencing approach, low coverage in CpG poor regions, no distinction between 5mC and 5hmC	Meissner et al. (2005), Gu et al. (2011)

COBRA-seq	Library construction, bisulfite conversion, restriction enzyme digestion, enrichment of cut fragments with biotin labels, amplification and sequencing	M	Yes	0.1–1	10–15	<1 %	1–4 M	Allows for analysis of DNA methylation in other sequence contexts than CpG and might be an interesting method in organisms with very little methylation, relative quantification only, less dependent on methylation density compared to MeDIP/MBD-seq	Varinli et al. (2015)
Agilent SureSelect^{XT} methylation capture	Libraries are prepared from genomic DNA fragments, hybridized to the capture probes, eluted, bisulfite converted, and amplified before being sequenced by a NGS	U+M	Yes	3	50	2.8 %	3.7 M	Off the shelf product, captures all known promoters and upstream regions, uses the procedure used for exome sequencing with an additional bisulfite conversion step, relatively expensive, custom design possible	Borno et al. (2012), Walker et al. (2015)

(continued)

Table 1 (continued)

Technology	The technology behind	*M*ethylated/ *u*nmethylated fraction	Single-nucleotide resolution (yes/no)	DNA input in µg	Required number of reads per sample in M	Coverage of the human genome	Coverage of CpGs in the human genome	Comments	References
SeqCap Epi CpGiant enrichment	Libraries are prepared from genomic DNA fragments, bisulfite converted, pre-amplified hybridized to the capture probes, eluted, and amplified before being sequenced by a NGS	U+M	Yes	1	40–50	2.7 %	5.5 M	Less input compared to the SureSelect capture as capture is performed on bisulfite-treated libraries after pre-amplification, relatively expensive, custom design possible	Walker et al. (2015), Li et al. (2015c)
MBD-seq/ MethylCap-seq	Enrichment of methylated DNA using a methyl-binding domain protein before or after a library construction followed by sequencing	M	No	0.2–1	30–40	95 %	17 M	Also suitable for FFPE, amenable to high-throughput and analysis of large cohorts, absolute quantification of DNA methylation difficult, specific for 5mC	Brinkman et al. (2010), Serre et al. (2010)

MeDIP-seq	Library construction, enrichment of methylated DNA using an antibody against 5mC followed by sequencing	M	No	0.3–0.5	50–60	95 %	17 M	Captures and allows to analyze also the transposons of the human genome, specific for 5mC, absolute quantification of DNA methylation difficult, lower signal to noise compared to MBD- and MIRA-seq, better coverage in CpG poor regions	Weber et al. (2005), Taiwo et al. (2012)
MIRA-seq	Enrichment of methylated DNA using a methyl-binding domain protein before or after a library construction followed by sequencing	M	No	0.1–1	40–50	95 %	17 M	Higher specificity compared to MBD-seq, absolute quantification of DNA methylation difficult, specific for 5mC	Jung et al. (2015)

(continued)

Table 1 (continued)

Technology	The technology behind	*M*ethylated/*u*nmethylated fraction	Single-nucleotide resolution (yes/no)	DNA input in µg	Required number of reads per sample in M	Coverage of the human genome	Coverage of CpGs in the human genome	Comments	References
MRE-seq	Digestion of genomic DNA in parallel with several methylation-specific restriction enzymes, size selection, library construction, sequencing	U	Yes, at restriction site	1.5–2.5	30	4 %	1.7 M	Analyzes preferentially unmethylated CpGs in high-density CpG regions, in combination with MeDIP comprehensive genome coverage is achieved	Maunakea et al. (2010), Li et al. (2015a)
MSCC/MSDK	Create sequence tags next to unmethylated restriction sites by adding a linker with type IIS restriction enzyme	U	Yes	20	50	<1 %	6 M	Requires large amounts of DNA, coverage can be increased using multiple restriction enzymes, relatively complicated workflow	Ball et al. (2009), Li et al. (2009)

HELP	Digestion of genomic DNA in parallel with a methylation-specific restriction enzymes and its methylation-insensitive isoschizomer, size selection, library construction, sequencing	U+M	Yes, at restriction site	0.3	30	4 %	3 M	Normalization of the counts signals from the two digestions allows for quantitative analysis at restriction sites, limited coverage due to the use of restriction enzymes, variation allows for the parallel analysis of 5hmC	Khulan et al. (2006)
DREAM-seq	Sequential digestion with the SmaI/XmaI isoschizomer pair, creates fragments with methylation-specific ends, adaptor ligation and sequencing	U+M	Yes, at restriction site	1–2	25	1.8 %	50,000	Analyzes preferentially unmethylated CpGs in high-density CpG regions (e.g., CpG islands), low number of CpGs covered	Jelinek et al. (2012)

(continued)

Table 1 (continued)

Technology	The technology behind	*M*ethylated/ *u*nmethylated fraction	Single-nucleotide resolution (yes/no)	DNA input in µg	Required number of reads per sample in M	Coverage of the human genome	Coverage of CpGs in the human genome	Comments	References
Infinium BeadChips	Genotyping of bisulfite-converted DNA	U+M	Yes	0.5	Microarray	2–3 %	480,000/ 840,000	Absolute quantification, cost-effective solution, amenable to the analysis of large cohorts, high data quality, standard bioinformatic pipelines available, limited coverage	Sandoval et al. (2011), Moran et al. (2016)

Abbreviations: *5mC* 5-methylcytosine, *5hmC* 5-hydroxymethylcytosine, *WGBS* *w*hole-*g*enome *b*isulfite *s*equencing, *RRBS* *r*educed *r*epresentation *b*isulfite *s*equencing, *COBRA* *co*mbined *b*isulfite *r*estriction *a*nalysis, *MBD* methyl-binding domain, *MeDIP* *me*thylated *D*NA *i*mmuno*p*recipitation, *MIRA* methylated-CpG island recovery assay, *MRE* *m*ethylation-specific *r*estriction *e*nzyme sequencing, *MSCC* *m*ethylation-*s*pecific *c*ut *c*ounting, *MSDK* *m*ethylation-specific *d*igital *k*aryotyping, *HELP* HpaII tiny fragment *e*nrichment by *l*igation-mediated *P*CR, *DREAM* *d*igital *r*estriction *e*nzyme *a*nalysis of *m*ethylation

Furthermore, the possibility to investigate regions of particular interest will depend on the presence of nearby restriction sites. Increased coverage of the genome can be obtained using multiple restriction enzymes for the initial complexity reduction step (Wang et al. 2013a; Martinez-Arguelles et al. 2014).

Similar to the originally devised technology for locus-specific DNA methylation analysis (Xiong and Laird 1997), *c*ombined *b*isulfite *r*estriction *a*nalysis (COBRA)-seq uses restriction enzymes to digest a bisulfite-converted sequence (Varinli et al. 2015). Therefore, in the COBRA assay, in contrast to the below-described methods using methylation-specific restriction enzymes, digestion does not directly depend on the presence of a methyl group but rather on the sequence obtained after bisulfite treatment, which is of course dependent on the original methylation state. After adaptor ligation and bisulfite treatment, fragments are amplified by limited PCR or linear amplification to avoid the presence of uracil in the sequence fragments, which could impede the subsequent restriction digestion. Fragments are digested with a restriction enzyme having a CpG dinucleotide in its recognition sequence. Adaptors are ligated, which have the complementary sequence at their 3′end corresponding to the recognition sequence. Biotinylation of the 5′ adaptor allows for the enrichment of cut molecules. After amplification by PCR or linear amplification using a T7 promoter included in the 5′ adaptor, samples are sequenced in multiplex on a HiSeq instrument (Varinli et al. 2015). Compared to the below-described methods (Sect. 6.2) using enrichment of the methylated fraction for the genome by protein affinity or antibody-based approaches, COBRA-seq might be less strongly influenced by methylation density, and the use of appropriate restriction enzymes allows also for the analysis of non-CpG methylation. This method, however, has a lower coverage due to its dependence on restriction enzyme recognition sites and, as all bisulfite methods, will not be able to distinguish 5mC from 5hmC. Furthermore, COBRA-seq will not yield absolute DNA methylation levels, but is better suited to identify differential DNA methylation between groups of samples.

Another method to enrich for methylated cytosines in bisulfite-converted DNA is *s*treptavidin *b*isulfite *li*gand *m*ethylation *e*nrichment (SuBLiME), for which, after complexity reduction using a restriction enzyme and bisulfite conversion, biotinylated dC is incorporated in a PCR reaction, which allows for the subsequent enrichment of the adaptor-ligated fragments, which originally contained methylated cytosines and correspond to the methylated fraction of the genome (Ross et al. 2013).

An alternative approach for a large-scale methylome sequencing uses long probes to capture bisulfite-treated DNA in regions of interest. This capture can be performed using either oligonucleotide microarrays (Hodges et al. 2009) or solution-based hybridization (Lee et al. 2011), and the capture can be performed on the bisulfite-converted fragments (Hodges et al. 2009) or prior to bisulfite conversion (Lee et al. 2011). If the capture is performed prior to conversion, this method requires a large amount of input DNA, while little DNA is recovered after hybridization, necessitating a high number of PCR cycles after bisulfite treatment. Furthermore, the harsh conditions of the bisulfite treatment can further reduce the complexity of the captured fragments. If capture is performed after bisulfite

conversion, this requires the design of probes complementary to all possible alleles generated by the bisulfite conversion to avoid preferential capture of molecules with a distinct DNA methylation pattern. While capturing pre-amplified bisulfite-converted DNA on a microarray (Hodges et al. 2009) can partially circumvent the problem of the required starting material, this approach leads nonetheless to a reduced percentage of the "on-target" sequences compared to the capture after conversion-based methods (~10 vs. 80 %). Solution-based hybridization methods present clear advantages in terms of flexibility and specificity. They make use of biotinylated RNA baits to capture regions of interest (Lee et al. 2011), similarly as in standard exome sequencing for the detection of genetic variations (Gnirke et al. 2009). Capture can be performed using products designed for genetic analysis, such as the SureSelect Human All Exon kit, prior to bisulfite conversion and DNA methylation analysis by sequencing (Wang et al. 2011b); however, for most scientific questions, these products contain only a small number of loci that might show variability in DNA methylation patterns.

Predesigned products enabling the capture of promoters and CpG islands are nowadays commercially available, such as Agilent's SureSelect™ Human Methyl-Seq covering ~3 % of the genome (84 Mb), corresponding to ~1/7 of all sequencing accessible CpG sites (3.7 M CpG sites (Borno et al. 2012)). The technology also allows customized capture panels targeting regions of interest (Ivanov et al. 2013). Although the required amount of starting material has been significantly reduced from 20 to 30 μg (Lee et al. 2011) to 2–3 μg (Walker et al. 2015) since the method was originally devised, this point remains one of the major drawbacks of these methods. This requirement is mainly due to the fact that the capture is carried out after adapter ligation and prior to the bisulfite conversion proscribing an amplification step. Furthermore, the high number of PCR cycles that need to be performed following the release of the captured fragments potentially distorts the DNA methylation values. To reduce the number of PCR amplification cycles after bisulfite treatment of the released DNA fragments, DNA from several hybridizations can be pooled prior to bisulfite conversion (Lee et al. 2011). However, recent advances show the feasibility of a post-bisulfite adaptor tagging (PBAT) protocol, reducing significantly the amount of required material (Miura and Ito 2015). Design of custom capture arrays targeting a high number of specific regions of interest makes this approach competitive to RRBS, which costs substantially less, but might also be less suited if specific regions that are mandatory for the analysis are not easily accessible by a (combination of) restriction digests. The SeqCap Epi CpGiant from Roche/NimbleGen allows for a reduced input of DNA (500 ng–1 μg) compared to the SureSelect Methyl-Seq system and is performed on a whole-genome bisulfite sequencing (WGBS) library by using capture probes complementary to methylated, unmethylated, and partially methylated targets after bisulfite conversion (Walker et al. 2015; Allum et al. 2015; Li et al. 2015c). Up to 5.5 M CpGs in ~80 Mb of the human genome can be interrogated simultaneously at a single-nucleotide resolution. A sophisticated design in combination with the use of long probes allows for a very efficient capture and focuses on only regions of interest at much increased coverage compared to WGBS (Allum et al. 2015). The probe design does allow pre-amplifying the sequencing

library prior to capture, and thereby starting from smaller amounts of input material compared to other capture approaches is possible. Up to four samples can be sequenced on a single lane of a flow cell of an Illumina HiSeq 2000 requiring 30 % less sequencing depth compared to the Methyl-Seq capture (70–80 M reads compared to 100–150 M reads). Read numbers as low as 40–50 M have been reported to be sufficient for a coverage allowing accurate quantitative assessment.

Several technologies have been developed that allow the rapid generation of a large number of amplification products simultaneously. Several thousand target regions can be amplified using the RainDance technology in microdroplets to analyze 10,000′s of CpGs (Paul et al. 2014; Komori et al. 2011). Up to 4000 target loci can be amplified from bisulfite-converted DNA simultaneously. DNA and PCR reagents, including a PCR primer pair for a specific region, are fused into picoliter-sized microdroplets, performing effectively a single-plex emulsion PCR in each droplet, thereby avoiding potential problems of multiplex PCRs. This approach has shown to provide excellent coverage of the targeted regions of up to 99 % (Komori et al. 2011), and the data was highly correlated with data obtained on Infinium Methylation 450 K BeadChips (Paul et al. 2014). Microdroplet PCR is also suitable for the analysis of FFPE samples, albeit at the price of reduced accuracy and increased dropout rates (call rate ~50 %) (Paul et al. 2014). An alternative approach uses padlock probes, horseshoe-shaped oligonucleotides with their both ends complementary to a bisulfite-converted sequence surrounding a CpG of interest (Deng et al. 2009; Ball et al. 2009). The region complementary to the target is filled in by a polymerase, and the padlock probe is circularized, which protects it against the exonuclease digestion used to remove linear DNA. The target regions are PCR amplified using primers with molecular barcodes and sequencing platform-specific adaptors annealing to the common backbone of the padlocks and subsequently sequenced. Similarly to the microdroplet PCR, this approach allows analyzing tens to hundreds of thousands of CpGs (Deng et al. 2009; Ball et al. 2009; Diep et al. 2012). As the primer design and setup of both padlock probes and microdroplet PCR requires significant upfront investment, as well as expertise in the design and some optimization, these technologies are most suited for the analysis of DNA methylation patterns of a large number of loci in either large cohorts or in a routine setting with a fixed panel of target regions.

6.2 Affinity- and Antibody-Based Enrichment Methods

In contrast to the above-described sequencing assays, several methods have been developed, which yield enrichment values for the methylation states over genomic regions but do not provide quantitative values of DNA methylation levels at individual CpGs. Genome-wide DNA methylation profiles with such a region-specific resolution can be obtained using methylated DNA immunoprecipitation (MeDIP) (Weber et al. 2005) or enrichment with methylated CpG binding proteins (Serre et al. 2010; Brinkman et al. 2010), which will yield alterations at the level of regions but not at single nucleotides. Several proteins are commonly used for MBD

sequencing approaches, including MBD2 and MECP2, and the inherent principle of affinity purification of methylated DNA is known under several names: MBD-isolated genomic sequencing (Serre et al. 2010), MCip (Gebhard et al. 2006), MBD-seq (Li et al. 2010a), MethylCap-seq (Brinkman et al. 2010), or MIRA-seq (Jung et al. 2015) (Table 1). Although MBDs might display a certain target specificity and MBD columns enrich methylated DNA significantly, they do not fully purify methylated sequences (Selker et al. 2003). These technologies provide a compromise for DNA methylation studies in large cohorts at a reasonable cost, which the methods covering all CpGs at single-nucleotide resolution are not yet able to deliver. Furthermore, these approaches are specific for the analysis of 5-methylcytosine and therefore allow distinguishing 5-methylcytosine from 5-hydroxymethylcytosine (Williams et al. 2011; Wu et al. 2011). The latter can be analyzed using specific antibodies against 5-hydroxymethylcytosine, as described in the section on methods for the analysis of 5-hydroxymethylcytosine (Sect. 15.10). The shift to NGS as a readout platform has significantly increased the accuracy, as the microarray-based protocols required a whole-genome amplification step, which led to a substantial bias in CpG-rich regions (Robinson et al. 2010). Nonetheless, it should not be forgotten that standard amplifications using, e.g., the Illumina adaptors are tailored to a specific GC content and regions with very high or low GC content might still be subject to amplification bias.

One of the major advantages of MeDIP-seq is the simple protocol and its resemblance with widely used ChIP-seq protocols, as well as the possibility for automation (Taiwo et al. 2012). Briefly, genomic DNA is fragmented using ultrasound or acoustic shearing, and (unmethylated) sequencing platform-specific adaptors are ligated to the fragmented DNA. This step has to be performed prior to the immunoprecipitation as the immunoprecipitation yields single-stranded DNA, which is incompatible with standard library preparation protocols (Taiwo et al. 2012). Methylated DNA is immunoprecipitated with high affinity and specificity with an antibody against 5-methylcytosine, and most commercial suppliers provide a monoclonal antibody from the same cell line. Immunoprecipitated fragments are released from the beads, PCR amplified, and sequenced. Earlier protocols also included size selection steps prior to the sequencing; however, the loss of material induced by this step needs to be compensated by additional PCR cycles potentially increasing the percentage of duplicate reads, and with appropriate bioinformatic analysis, a size selection step can be omitted. Improved and automated protocols allow for the standardized high-throughput analysis of samples with little starting material (Taiwo et al. 2012). The number of sequencing reads in MeDIP-seq depends on both the DNA methylation level and the CpG content of the sequence. Computational algorithms are therefore required to account for this bias in amplification and to convert signal intensities into a methylation percentage (Down et al. 2008; Pelizzola et al. 2008; Huang et al. 2012; Xiao et al. 2015). MeDIP-seq is one of the most widely used genome-wide approaches for DNA methylation analysis (Yuan et al. 2014; Feber et al. 2011; Taiwo et al. 2013).

In combination with the complementary *m*ethylation-specific *r*estriction *e*nzyme sequencing (MRE-seq) approach, which makes use of methylation-specific

restriction enzymes to map unmethylated cytosines within restriction recognition sites at single-nucleotide resolution (see also Sect. 6.3), a comprehensive coverage of ~80 % of the CpGs in the human genome can be obtained (Li et al. 2015a).

MeDIP-seq does not provide data with single-nucleotide resolution, which needs to be obtained by locus-specific analyses such as (pyro)sequencing or mass spectrometric analysis of the regions of interest and thus requires extensive follow-up studies. Partial data on the DNA methylation patterns at single-nucleotide resolution can be obtained by bisulfite conversion of the immunoprecipitated methylated DNA fragments. However, as the fragments are enriched for completely methylated molecules (Sengenes et al. 2010), information of variable DNA methylation patterns will be lost, and the observed profiles might not be representative of the methylation states of the original sample. While mostly performed on Illumina sequencers, protocols have also been devised for the Ion Torrent platform (Corley et al. 2015).

In the second capture-based approach, the methylated fraction of a genome is isolated by affinity purification of methylated DNA with MBD proteins. Of the different MBD proteins, MBD2b has the highest affinity for methylated DNA (Fraga et al. 2003). For the affinity-based enrichment with MBD proteins, genomic DNA is sonicated prior to capture with the respective MBD protein coupled to a solid support, such as streptavidin beads. Following the capture reaction, the bound methylated DNA can be eluted as a single fraction or in several fractions using increasing concentrations of salt in the elution buffer, which enables to target fractions with a specific CpG density, because fragments with a high density of CpGs are eluting last. As double-stranded DNA is recovered from the elution, the library preparation can be performed after affinity purification. MBD-seq methods can also be applied to fragmented DNA, such as DNA extracted from FFPE tissues. It is amenable to very high throughput through automation, and the analysis of very large cohorts including 1500 samples (750 schizophrenia patients and 750 controls) has been reported (Aberg et al. 2012). Information at single-nucleotide resolution can be obtained by bisulfite conversion of the captured DNA and subsequent massively parallel sequencing (Brinkman et al. 2012).

In the *m*ethylated-CpG *i*sland *r*ecovery *a*ssay (MIRA), a glutathione S-transferase (GST)-tagged full-length MBD2b is used to bind sonicated methylated DNA fragments, and the affinity to methylated CpG dinucleotides is further enhanced in a dose-dependent manner by the addition of the MBD3-like-1 protein (Rauch and Pfeifer 2005). The combined effect significantly improves the sensitivity of the assay, and a single methylated CpG dinucleotide allows for capture of the corresponding DNA molecule. Ligation of oligonucleotide linkers to enzymatically digested DNA prior to affinity chromatography permits efficient amplification of eluted fractions and subsequent analysis of input DNA- and MIRA-enriched amplification products by NGS (Jung et al. 2015; Choi et al. 2010). Of note, as the capture is performed on double-stranded DNA, library preparation can be performed before or after the enrichment step. There are currently a number of commercial kits for MBD sequencing on the market that have recently been compared to each other, highlighting some differences in performance between the different kits (De Meyer et al. 2013; Aberg et al. 2015). As for MeDIP-seq, CpG density and GC content of

the fragmented DNA are major biases affecting the efficiency of the affinity purification, as well as the subsequent sequencing (Robinson et al. 2010). Therefore, computational approaches have been developed for background estimation, to correct for CpG coverage and CpG density in reads (Lan et al. 2011; Riebler et al. 2014).

Instead of using antibody or protein affinity purification, methylated cytosines can also be chemically modified after oxidation with TET enzymes, followed by the attachment of a biotin group (TAmC-seq) (Zhang et al. 2013). After protection of 5hmC with a glucose moiety (as described in Sect. 15.10 on the analysis of 5-hydroxymethylation) (Song et al. 2011), methylated cytosines are oxidized to hydroxymethylated cytosines using a catalytically active fragment of the TET enzyme, modified with a glucose molecule with an azide group, which by "click" chemistry can be used for the labeling with biotin or other chemical tags. This label can subsequently be used to enrich selectively for sequences with methylated cytosines. Although this method has been demonstrated to have less density bias and allow a more even coverage, it has been little used probably due to its multi-step protocol and the fact that some of the reagents might not be readily available in all laboratories.

Both MBD-seq and MeDIP-seq are well suited for the identification of differentially methylated regions, and when correctly accounted for CpG density, they can distinguish between methylated and unmethylated regions with a precision similar to RRBS (Bock et al. 2010; Harris et al. 2010), Infinium BeadChips (Harris et al. 2010; Clark et al. 2012), or the mass spectrometry-based EpiTYPER assay (Nair et al. 2011). The quantitative accuracy is nonetheless reduced for regions with intermediate DNA methylation levels (Harris et al. 2010), and despite the advantage of having a higher coverage of the genome compared to RRBS, substantially more sequencing is required (40–60 M reads for MBD/MeDIP vs. 20 M reads for RRBS) (Bock et al. 2010). Both technologies suffer from false-positive signals in repetitive CpG-rich regions, in which minor methylation differences are amplified through the enrichment, while copy number variations do generally not seem to influence the DNA methylation profile (Bock et al. 2010; Robinson et al. 2010). There are also some differences in the target distribution between MeDIP and MBD (Nair et al. 2011). While MBD-seq enriches preferentially CpG islands and regions with high CpG density, MeDIP does also enrich for methylated CpGs in regions with low CpG density, and in general the signal over baseline ratio is lower for MeDIP (Bock et al. 2010; Nair et al. 2011; Robinson et al. 2010). In practice this means that MeDIP will require a significantly higher number of reads to capture all methylated cytosines compared to MBD-seq and might identify less DMRs at the same read depth. Nonetheless, MeDIP was found to display the best coverage for the whole genome and especially for gene body regions (Walker et al. 2015). Furthermore, while MeDIP can detect methylated cytosines in any sequence context, the MBD-based approaches will only detect CpG methylation. If the target organism might contain methylated cytosines in other sequence contexts, such as in plants, MeDIP will allow for a more comprehensive DNA methylation profiling. Of note, the CpG density of the DNA fragments eluted for sequencing using the MBD-seq approaches

can be modulated by using different salt concentrations for elution (Serre et al. 2010). The absence of methylation is difficult to assess using these enrichment methods as a lack of reads in a given region can be due to the absence of methylation but could also be due to technical problems, such as lack of amplification or just chance.

Complementary information on unmethylated cytosines and regions of the genome can be obtained using an enrichment technology that has been termed mTAG-seq (Kriukiene et al. 2013). In this method, unmodified cytosines are chemically modified using an engineered M.SssI methyltransferase and synthetic AdoMet analogs, followed by tagging of the AdoMet analog with a covalent biotin molecule, which is subsequently used for enrichment and sequencing.

6.3 Sequencing Approaches Using Methylation-specific/Methylation-Dependent Restriction Enzymes

Methylation-specific restriction endonucleases, i.e., enzymes that are blocked by methylated cytosines in their recognition sequence (Bird and Southern 1978), have been widely used in the past for the analysis of methylation patterns in combination with their methylation-insensitive isoschizomers. There are about 50 methylation-specific restriction enzymes known, but few of them are available in combination with a methylation-insensitive isoschizomer. One of the most commonly used pairs of enzymes is HpaII/MspI; both recognize and cleave the four-base palindrome C^CGG in double-stranded DNA, but while MspI cleaves the DNA independent of the methylation status of the central CpG site, HpaII is unable to cleave when the second cytosine is methylated ($C^{\wedge me}CGG$). Although methods based on methylation-specific restriction enzymes are simple and relatively cost-effective as they do not require any special instrumentation, they are hampered by the limitation to specific restriction sites, as only CpG sites found within these sequences can be analyzed. For example, only ~4 % of the CpG sites in non-repetitive sequences are located in HpaII recognition sites, and only 0.03 % can be cleaved by NotI (Fazzari and Greally 2004). The fraction of the genome interrogated is often arbitrary and not associated with a specific functionality. In addition, methods using these enzymes might be prone to false-positive results due to incomplete cleavage, and some sequences are intrinsically resistant to digestion if not appropriately controlled. For example, non-CpG methylation on cytosines or DNA adducts in the vicinity of the cleavage site might influence the restriction capacity of an enzyme. Digestions are therefore difficult to perform on material extracted from formaldehyde-fixed paraffin-embedded (FFPE) samples. Although in many cases restriction enzyme-based approaches have been replaced by more quantitative sequencing methods allowing a more rapid identification of altered DNA methylation levels at higher spatial resolution, a few protocols are still under use and have been successfully transferred to NGS instruments, which permit a more quantitative analysis of the isolated (methylated or unmethylated) fraction of the genome (Table 1).

*M*ethylation-specific *r*estriction *e*nzyme sequencing (MRE-seq) identifies unmethylated CpG sites at single CpG site resolution by sequencing size-selected

fragments from parallel DNA digestions with a number of methylation-specific restriction enzymes (e.g., HpaII, Hin6I, and AciI) and covers ~ 1.7 M of the 28 M CpG sites in the human genome (Maunakea et al. 2010). After restriction digest and size selection, Illumina adaptors are ligated to the DNA fragments, PCR amplified, and sequenced using short read (50 bp) single-end sequencing, thereby identifying unmethylated CpG sites within the restriction sites with single-base resolution. Interrogation of the methylated fraction of the same sample by MeDIP-seq, described in more detail in the previous section, yields complementary information, and the combined use of MRE-seq and MeDIP-seq allows for genome-wide DNA methylation analysis at high coverage and resolution while limiting the biases of each technology (Li et al. 2015a). The combined approach analyzes up to 22 M of the 28 M CpGs of a human genome, and the quantitative accuracy of the methylation levels obtained by MeDIP in regions with low or intermediate DNA methylation is improved (Harris et al. 2010). Similarly, the methylated fraction of the genome can also be interrogated using methylation-dependent restriction enzymes of the mrr-like family (FspEI, MspJI, LpnPI, AspBHI, etc.) used in the MethylRAD assay, which creates small fragments of 32 nucleotides around a methylated cytosine that are subsequently sequenced (Wang et al. 2015b). A very similar protocol to MRE-seq, termed Methyl-MAPS, isolates both the methylated and unmethylated fraction of the genome by using the methylation-dependent restriction endonuclease McrBC and a combination of several restriction enzymes, respectively, prior to library preparation and paired-end sequencing (Edwards et al. 2010).

(Modified) *m*ethylation-*s*pecific *d*igital karyotyping ((M)MSDK) and *m*ethylation-*s*pecific *c*ut *c*ounting (MSCC) are similar to MRE-seq in that a methyl-sensitive restriction enzyme is employed but includes additional steps that reduce the amount of sequencing required (Li et al. 2009; Ball et al. 2009). Rather than sequencing the unmethylated and methylated fractions of the genome, specific regions of the genome can be identified from short sequence tags in a strategy similar to the serial analysis of gene expression (SAGE) approach used for RNA expression analysis. After an initial digestion with HpaII, adaptors with a type II restriction enzyme (cutting outside its own recognition site) recognition site, such as MmeI, are ligated, which allows generating small sequence fragments (20–30 nucleotides), which are subsequently sequenced and mapped back to the genome. This strategy significantly reduces the amount of sequencing and in turn reduces the costs of this approach. Increased coverage can be obtained by using a cocktail of different methylation-specific restriction enzymes for the initial digest (Colaneri et al. 2011). Analysis of the methylated fraction of the genome with a similar strategy is possible using methylation-dependent type IIS restriction enzymes generating 32 nucleotide sequences for downstream analysis (Cohen-Karni et al. 2011).

In the initial protocol of *d*ifferential *m*ethylation *h*ybridization (DMH) (Huang et al. 1999), two differentially treated fractions of the same sample were comparatively hybridized to an array-based probe library, and relative (fluorescent) intensities were measured. Genomic DNA was fragmented by digestion with a frequently cutting restriction enzyme that preferentially cuts outside of the CpG islands. Linkers for PCR amplification were subsequently ligated to the digestion products. The sample was then split into two parts and one half digested with a methylation-specific

restriction enzyme. In this sample, only methylated fragments that are resistant to the digestion are amplified in the subsequent PCR, while in the reference sample all fragments are amplified. Fractions are labeled with two different fluorescent dyes and hybridized to microarrays with an immobilized CpG island library. The HELP (*H*paII tiny fragment *e*nrichment by *l*igation-mediated *P*CR; Khulan et al. 2006) assay or the very similar Methyl-Seq protocol (Brunner et al. 2009) use a digestion of the same sample with MspI, the methylation-insensitive isoschizomer of HpaII as a reference, thereby allowing to normalize the signal for the HpaII digest by that of the MspI digest. This assay format has been successfully transferred to sequencing platforms and has been used in a number of studies (see, e.g., Yuan et al. 2016, Hu et al. 2014). The HELP assay allows analyzing quantitatively ~10 % of the CpGs, thereby covering a similar percentage of CpGs compared to RRBS (Khulan et al. 2006; Suzuki et al. 2010). Of note, a variation of the protocol does also permit a separate detection of 5-hydroxymethylcytosine (Bhattacharyya et al. 2013).

*D*igital *r*estriction *e*nzyme *a*nalysis of *m*ethylation (DREAM) is based on sequential cuts of the genomic DNA with a pair of restriction enzymes (SmaI and XmaI) at CCCGGG target sites (Jelinek et al. 2012). Unmethylated sites are first digested with SmaI, which cuts unmethylated CCCGGG sites in the middle (at CCC^GGG), leaving behind blunt-ended fragments. In the next step, the remaining methylated sites are digested with XmaI, which is not blocked by CpG methylation. It cuts the recognition site sideways at C^CCGGG, forming 5′-CCGG overhangs. The sequential cuts thus create distinct methylation-specific signatures at the ends of restriction fragments: 5′-GGG for unmethylated CpG sites and 5′-CCGGG for methylated sites. The DNA fragments resulting from the digestions are ligated to barcoded NGS adapters; libraries with distinct barcodes are pooled and sequenced using a paired-end protocol. The sequencing reads are aligned to the genome and mapped to unique CCCGGG target sites. Methylation at individual CpG sites is calculated as the digital counting of sequencing reads with the methylated signature to the total number of reads mapping to the site. Quantitative accuracy can be improved using spike-ins with defined levels of DNA methylation at the enzyme recognition sites. Sequencing of 25 million reads per sample typically yields accurate determination of DNA methylation levels at 50,000 unique CpG sites with high coverage (Jelinek et al. 2012).

For the analysis of gene-specific methylation patterns or individual CpG positions, methods using methylation-specific endonucleases have largely been replaced by PCR-based methods following treatment of genomic DNA with sodium bisulfite, although as mentioned briefly in Sect. 15.7 they are still used and do provide some advantages when combined with high-throughput qPCR and multiplex amplification systems.

6.4 Epigenotyping Arrays

With whole-genome bisulfite sequencing being not yet affordable at a large scale and the low resolution of antibody and methyl-binding protein enrichment of methylated DNA, epigenotyping technologies have emerged as an alternative tool for

the identification of differentially methylated regions and DNA methylation-based biomarkers. Epigenotyping technologies, such as the Infinium Human Methylation 450 K or EPIC BeadChip (Illumina Inc., CA, USA), generate a methylation state-specific “pseudo-SNP” through bisulfite conversion, thereby translating differences in DNA methylation patterns into sequence differences that can be analyzed using quantitative genotyping methods (Bibikova et al. 2009, 2011). The 450 K BeadChip has dramatically expanded the genome coverage compared to previous generations of the BeadChip, analyzing more than 480,000 CpG sites covering 99 % of all RefSeq genes with an average of 17 probes per gene. This array is not only focused on CpG islands, but probes are distributed over various functional elements that are more prone to alter their DNA methylation status in response to environmental conditions or in cancer, such as CpG island shores and shelves (Irizarry et al. 2009). These arrays have been widely used for large-scale high-throughput studies as they employ highly standardized protocols that can be implemented with a large degree of automation into existing genotyping pipelines. The analysis of the results is relatively straightforward when compared to the required correction for CpG density or the cost- and time-intensive bioinformatic calculations needed in sequencing-based DNA methylation analysis, but it requires specific normalization protocols due to the combination of two different assay chemistries on the 450 K BeadChip that display a different dynamic behavior (Touleimat and Tost 2012; Aryee et al. 2014; Morris et al. 2014): The Infinium I probes (InfI), which represent ~1/3 of the array, convey information about the methylation state in the type of the bead (InfI). Two different bead codes are used to interrogate allele-specifically the base following an unmethylated or methylated cytosine (T or C after bisulfite conversion) at the same genomic location. If the 3′end of the probe hybridizes correctly, the probe is extended with the bases following the potential methylation variable position. The extended base is thus the same for methylated or unmethylated alleles, which means that the fluorescent signal does not carry any information on the methylation status. The Infinium II (InfII) probes are attached to a single type of beads, and the methylation information is obtained through dual channel single-nucleotide primer extension with labeled dideoxynucleotides on the methylation variable position of a CpG. The use of the two chemistries is a compromise to ensure the coverage of a large number of CpGs throughout the entire genome and also in CpG island-associated promoter regions. Because the InfII probes require a single probe to interrogate CpG positions, the number of CpGs potentially analyzed on the BeadArray is increased. However, they can only tolerate up to three CpG positions in the 50-mer probe, which are covered using degenerate bases, e.g., R (A/G), complementary to the potentially methylated position after bisulfite treatment. The InfI probes tolerate more variable position, but it should be noted that their design assumes an identical methylation pattern (methylated or unmethylated) of CpGs underlying the 50-mer probe sequence. Therefore, InfI probes are more adequate to analyze CpG positions in regions of high CpG density such as CpG islands. InfI/InfII is therefore not equally distributed among functional or CpG island-based categories present on the array.

Although data from the BeadChips is relatively easy to generate, a number of technical issues, such as batch effects, or underlying SNPs might confound the analysis if not appropriately controlled for. Furthermore, the low sequence complexity of the bisulfite-converted DNA could induce cross-hybridization events, and a number of potentially problematic probes have been reported (Chen et al. 2013). However, when properly handled, these epigenotyping arrays are accurate and display high correlation to RRBS (Bock et al. 2010) as well as locus-specific quantitative assays such as pyrosequencing (Roessler et al. 2012) or MethyLight (Campan et al. 2011).

The BeadChips can be applied to a variety of biological specimens including fresh frozen samples, as well as formalin-fixed paraffin-embedded (FFPE) samples or PAXgene-conserved samples. While DNA extracted from fresh frozen and FACS- or MACS-sorted cells can be directly used on the array, FFPE samples, which are the most common form of tissue preservation in pathological archives, are not suitable for the direct use in the procedure due to the extensive cross-linking, fragmentation, and generation of apurinic/apyrimidinic sites, which all impede enzymatic processing steps such as whole-genome amplification, which is an essential step in the 450 K protocol. A restoration method for FFPE DNA uses a ligation-based approach to obtain DNA fragments of sufficient size (Thirlwell et al. 2010). However, there is still a controversy about the concordance between differentially methylated loci detected in fresh frozen tissue and DNA restored from FFPE tissue (Jasmine et al. 2012). This approach has been found in two recent publications to yield relatively good results, leading to similar biological findings to those obtained from fresh frozen samples (Moran et al. 2014; Dumenil et al. 2014). However, it was also pointed out that there are substantial differences, which prohibit a combined analysis of fresh frozen and FFPE in the same study. PAXgene-preserved samples provide a new source for the analysis of a wide range of biomolecules and have shown to provide superior results compared to restored FFPE samples (Andersen et al. 2014).

Nonetheless, these arrays do analyze only a small number of the 28 million CpG sites of the human genome, and no commercial arrays for the analysis of nonhuman samples are currently available. As all bisulfite-based analysis techniques, they are not able to differentiate between cytosine methylation and hydroxymethylation. Specialized protocols based on oxidative bisulfite conversion or TET-assisted bisulfite analysis allow for the assessment of hydroxymethylation also on the 450 K BeadChip platform (Stewart et al. 2015; Nazor et al. 2014).

The 450 K array provides currently a good compromise between coverage, throughput, cost, resolution, and accuracy, permitting genome-wide epigenome analysis by epigenotyping, and has been rapidly adapted by the community for epigenome-wide association studies for the analysis of a large variety of diseases and phenotypes (see, e.g., Miceli-Richard et al. 2016; Grundberg et al. 2013; Dick et al. 2014; Glossop et al. 2014; Martino et al. 2014; Stefansson et al. 2015).

Very recently the Illumina MethylationEPIC BeadChip was presented and evaluated, which adds ~ 400,000 CpGs to the content of the Illumina HM450 array, focusing essentially on enhancer regions identified in the ENCODE and FANTOM5

projects, thereby extending significantly the information content of the BeadChip (Moran et al. 2016). The addition of the intergenic, gene regulatory regions, which do often display intermediate and variable DNA methylation levels, makes this array perhaps also more suited for the analysis of hydroxymethylcytosine.

7 Locus-Specific DNA Methylation Analysis

While all the above-described technologies are well suited for the identification of differentially methylated genes, they are still too costly and/or do not provide the required analytical sensitivity and specificity for detailed locus-specific analyses. Large DNA methylation data sets from clinical samples are now available and can be mined for clinical associations in public databases or through the TCGA/ICGC project portals. However, identified DNA methylation-based biomarkers need to be validated and replicated using locus-specific methods for DNA methylation analysis. Furthermore, technologies for the potential use in a clinical setting have to be cost-effective, sensitive, and specific. It would be a non-negligible advantage if the method can also be applied to DNA extracted from formalin-fixed paraffin-embedded (FFPE) clinical specimens. Furthermore, it would be preferable if the analysis can be performed at a high-throughput scale and in a relatively short time span. In general, closed-tube assay formats should be preferred to avoid cross contamination and thereby false-positive results. Following the interest in DNA methylation for the various clinical applications, many technologies have been developed for the quantitative analysis of DNA methylation patterns or levels at specific loci, mostly relying on the conversion of genomic DNA with sodium bisulfite (Table 2) (Tost 2008b; Laird 2003). While a large number of methods have been devised, relatively few methods are currently in use and are amenable to the throughput that is nowadays required. In this chapter, only the most commonly used methods including amplicon bisulfite sequencing, pyrosequencing, mass spectrometry, and real-time methylation-specific PCRs are described in detail. When single-nucleotide resolution is not required, other methods such as *m*ethylation-*s*pecific *h*igh-*r*esolution *m*elting analysis (MS-HRM) (Wojdacz et al. 2008b) or *m*ethylation-*s*pecific *m*ultiplexed *l*igation *p*robe *a*mplification (MS-MLPA) (Nygren et al. 2005) might be useful, for example, in case of screening for differentially methylated regions or if a method not relying on bisulfite treatment is required for validation (for MS-MLPA). MS-HRM makes use of the melting profiles of PCR products for a target locus originating from methylated and unmethylated variants of the same template, which differ after bisulfite treatment in their GC content. Therefore, the methylation status of an unknown sample can be determined by comparing the melting profile of the sample to calibration standards. A gradual increase of the temperature leads to a stepwise dissociation of the double strand in domains of the PCR product in function of their GC content differing between methylated and unmethylated molecules after bisulfite treatment. The application of this simple and cost-efficient technology, devised a long time ago (Worm et al. 2001), was previously limited by the toxicity of the intercalating agent *SYBR Green I* for DNA polymerases,

Table 2 Technologies for the analysis of locus-specific DNA methylation patterns

Technology	The technology behind	Advantages	Inconveniences	Throughput	Limit of detection or quantitative resolution	Suitability for cell-free DNA methylation analysis	References
Sanger bisulfite sequencing	Bisulfite conversion, PCR amplification, and in most cases cloning of PCR products. Sanger sequencing of clones and/or bulk amplification product	Standard equipment, single-nucleotide resolution, clonal methylation patterns (if cloned fragments are analyzed)	Quantitative resolution is usually low, high workload, and cost-intensive when combined with cloning	+/++	~10–15 %	–	Zhang et al. (2009)
Amplicon BS-Seq	Regions of interest are PCR amplified from bisulfite-converted DNA and libraries are constructed and sequenced by NGS (MiSeq/IonTorrent)	Clonal fragments, high coverage, multiple amplicons and individuals can be analyzed in parallel, can be combined with microfluidics and/or droplets for PCR product preparation	Analysis requires some bioinformatic expertise, relatively expensive	+++	~1–5 %	++	Korbie et al. (2015), Masser et al. (2013), Paliwal et al. (2013)
Pyrosequencing	Bisulfite conversion, PCR amplification, and sequencing-by-synthesis method. Incorporation of a complementary nucleotide releases a phosphate group that will be involved in the ATP-driven oxidation of luciferin. The light emitted from this reaction is monitored	Highly quantitative, single CpG resolution	High workload and feasible only for a limited number of target regions	++	~5 %	±	Tost and Gut (2007)

(continued)

Table 2 (continued)

Technology	The technology behind	Advantages	Inconveniences	Throughput	Limit of detection or quantitative resolution	Suitability for cell-free DNA methylation analysis	References
MassCLEAVE/ EpiTYPER	T7-promoter-tagged bisulfite-specific PCR followed by ssRNA transcription and base-specific cleavage by RNase A. The mixture is analyzed quantitatively by MALDI-TOF MS	Allele-specific methylation can be addressed, amenable to high-throughput and identification of heterogeneous DNA methylation patterns	Not all CpG positions can be analyzed, requires expensive instrumentation	+++	~5 %	+	Ehrich et al. (2005), Coolen et al. (2007)
MS-HRM	High-resolution melting curve analysis following quantitative real-time PCR RT-PCR yields quantitative methylation levels. Inclusion of a few CpG sites in the 5′ end of MIP primers allows for the amplification of unmethylated as well as methylated DNA	Rapid screening for the presence of differential DNA methylation	A limited number of CpG sites between the primers are necessary to obtain highly distinguishable melting curves but lead to biased amplification. Heterogeneous DNA methylation patterns are difficult to resolve	++	~5–20 % (depending on primer design)	++	Wojdacz et al. (2008b)

MS-MLPA	MLPA primers containing an additional restriction site are hybridized to the target of interest. Methylation-specific endonucleases digest the probes hybridized to unmethylated targets. Thus, only methylated targets are amplified	There is opportunity for multiplexing. MS-MLPA resolves the copy number variations simultaneously with the methylation status. Does not rely on bisulfite conversion	Difficult primer design and assay optimization, depends on the presence of methylation-specific restriction enzymes	±	~10 %	±	Nygren et al. (2005)
MS-qPCR	Digestion of DNA with methylation-specific restriction enzymes and subsequent amplification of undigested target DNA monitored quantitatively in real time	Simple assay design, no special instrumentation, amenable to high-throughput using microfluidic devices	Depends on the presence of methylation-specific restriction enzymes, Incomplete digestion gives false-positive results	+++	~5 %	++	Oakes et al. (2006), Wielscher et al. (2015)
MS-SNuPE/ SNaPshot	The region of interest is amplified and labeled ddNTPs are used to extend a primer that terminates at the CpG site. The products are visualized on a gel or by capillary electrophoresis	Multiplexing assays enable investigation of several targets of interest	Only specific CpG positions are analyzed	–	~5 %	–	Gonzalgo and Jones (1997), Kaminsky et al. (2005)

(continued)

Table 2 (continued)

Technology	The technology behind	Advantages	Inconveniences	Throughput	Limit of detection or quantitative resolution	Suitability for cell-free DNA methylation analysis	References
COBRA	Restriction enzymes are used to digest PCR products amplified from bisulfite-treated DNA, which contain a sequence difference in the enzymatic recognition site in function of the methylation status prior to the bisulfite conversion	Easy to implement and possibility for quantification	Gel-based readout, targets may be missed due to inhibition of enzymatic activity	–	–10 %	–	Xiong and Laird (1997), Brena et al. (2006)
MSP-PCR	At least one CpG site at or near the 3′ of the amplification primers ensures amplification of methylated DNA only. By including non-CpG cytosines in the primer sequence and having stringent annealing temperatures, amplification of incomplete converted and nontarget DNA is limited	MSP is a cost-effective approach, detects one methylated molecule in background of 1000 molecules	Qualitative readout, high rate of false positives and negatives	++	1 %	++	Herman et al. (1996)
qMSP	Same as MSP, but amplification is followed in real time	Sensitive technology, more cost-efficient than MethyLight while better control over false-positive and false-negative results compared to MSP	Heterogeneous DNA methylation might lead to amplification failure and thus false-negative results	+++	1 %	++	Nikolaidis et al. (2012)

SMART-MSP	Same as qMSP followed by a melting curve analysis allowing for the exclusion of some false-positive results	Cost-effective, no probe required, and yields a low-rate of false positives	Heterogeneous DNA methylation might lead to amplification failure and thus false-negative results or difficult to interpret melting curves	+++	1 %	+++	Kristensen et al. (2008)
MethyLight	Cleavage of a fluorescent hydrolysis probe designed to hybridize between MSP primers, principle based on the TaqMan principle. Detects one methylated molecule in background of 10,000 molecules	Highly sensitive; a reference assay (Alu or β-Actin) permits the quantification of the methylation level. Further improved sensitivity in a digital	Heterogeneous DNA methylation might lead to amplification failure and thus false-negative results	+++	~0.1 %	+++	Eads et al. (2000), Yu et al. (2015b)
HeavyMethyl	Blockers, designed to bind unmethylated DNA only, ensure the amplification of methylated DNA with methylation-independent amplification primers. A probe hybridizing to CpG sites containing a fluorophore and a quencher is designed to detect the amplification of methylated molecules	Highly sensitive; principle used in FDA-approved DNA methylation assays	Sophisticated design requiring multiple probes for each assay	+++	~0.1	+++	Cottrell et al. (2004)

Abbreviations: *BS* bisulfite, *MS* methylation sensitive, *MS-HRM* *m*ethylation-*s*pecific *h*igh-*r*esolution *m*elting, *MS-MLPA* *m*ethylation-*s*pecific *m*ultiplexed *l*igation *p*robe *a*mplification, *MS-SNuPE* *m*ethylation-*s*pecific single-*nu*cleotide primer extension, *COBRA* *co*mbined bisulfite restriction analysis, *MSP* methylation-specific *P*CR, *SMART-MSP* *s*ensitive *m*elting *a*nalysis after *r*eal-time methylation-specific *P*CR

which prohibited working at the required saturating concentrations. Advances in fluorescence detection technology, new algorithms for data calculation, and the use of novel dyes permitted the development of this high-resolution melting analysis (HRM) (Wojdacz and Dobrovic 2007), allowing for the rapid scanning of a large number of genes for the presence of differential DNA methylation. Careful primer design (e.g., including or not including cytosine residues) allows to fine-tune the discriminative window to a specific range of DNA methylation levels, facilitating discrimination of samples with small differences (Wojdacz et al. 2009). MS-HRM can also be used on FFPE samples without a major loss of accuracy in the DNA methylation assessment (Daugaard et al. 2015). MS-HRM has been applied to the detection of aberrant methylation profiles in imprinting disorders (White et al. 2007; Alders et al. 2009; Wojdacz et al. 2008a), in cancer (Balic et al. 2009; Gupta et al. 2014), and in epidemiological studies analyzing environmental exposure (Li et al. 2015b, 2016). Of note, this technology does not provide DNA methylation profiles at single-nucleotide resolution, which will complicate its potential implementation in clinical diagnosis. However, optimized protocols have been devised for the subsequent pyrosequencing of the amplification products, which can then yield detailed information on the methylation status of the CpGs included in the amplification product (Candiloro et al. 2011).

MS-MLPA has been widely used for the diagnosis of imprinting disorders. While initially devised for the analysis of copy number alterations, this technique has been useful for the parallel analysis of up to 40 loci permitting a comprehensive analysis for all possible variations of DNA methylation aberrations in imprinting disorders (Dikow et al. 2007; Priolo et al. 2008; Henkhaus et al. 2012), the combined analysis of genetic and epigenetic alterations in imprinting disorders (Scott et al. 2008), as well as the tumor analysis (Serizawa et al. 2010; Homig-Holzel and Savola 2012). Two oligonucleotides with universal primer binding sites are annealed to a target region and are ligated in case of complete target complementarity. A methylation-specific enzyme is added to the ligation reaction digesting unmethylated templates and reducing the amount of ligated product. A semiquantitative readout is then performed using capillary electrophoresis allowing the detection of methylation differences of 10 % or more compared to the standards. While this technology does not rely on bisulfite conversion, it might yield false-positive results if the digestion is not complete and limits the applicability of the method to targets with restriction enzyme recognition sites.

PCR amplification following methylation-specific restriction digestion is an alternative strategy that requires substantially less DNA and no prior bisulfite conversion treatment and is well suited as a rapid screening tool for differential methylation (Singer-Sam et al. 1990). Multiple targets can be simultaneously analyzed by locus-specific multiplex PCR following methylation-specific restriction digest of genomic DNA (Melnikov et al. 2005). Additional information on the methylation status of a target region can be achieved by digesting the DNA with either methylation-specific restriction enzymes or methylation-dependent enzymes such as McrBC, thus allowing to distinguish complete methylation, partial methylation, or absence of methylation in the sequence (Yamada et al. 2004). Quantification can be

improved by monitoring the increase in fluorescence by quantitative real-time PCR with the intercalating dyes (Bastian et al. 2005; Oakes et al. 2006). Combined with microfluidic preparation of the PCR products, this method allows the analysis of a large number of target sequences from a very limited amount of starting DNA (Wielscher et al. 2015). To minimize false-positive results due to incomplete digestion, DNA is overdigested using a combination of several restriction enzymes, and two or better three restriction sites have to be present within the target sequence.

7.1 Amplicon Bisulfite Sequencing

Established methods, such as pyrosequencing (Tost and Gut 2007) or mass spectrometry-based DNA methylation analysis (Ehrich et al. 2005), which are described in more detail below, allow for the quantitative analysis of DNA methylation in a region of interest and are well suited for the analysis of a limited number of regions in a large number of samples, but they are difficult to upscale if a large number of potential candidate regions identified in genome-wide analyses have to be verified. The same is true for bisulfite Sanger sequencing that, coupled with PCR amplicon cloning, had been the "gold standard" technology for DNA methylation analysis in the past. In this method, the quantitative resolution depends on the number of clones that were analyzed, and this number was in most cases limited making results prone to artifacts (Grunau et al. 2001). Recently devised quantitative sequencing approaches make use of benchtop sequencers (e.g., Illumina's MiSeq; 454/Roche's Junior or Ion Torrent's PGM) and allow generating high levels of coverage (e.g., 100's–1000's ×) that yield precise measurements of the quantitative levels of cytosine methylation. Amplicon bisulfite sequencing using these instruments has become a widely used approach for the validation of genomic regions following methylome analyses and for answering hypothesis-driven research questions. In addition, due to the sequencing of clonal clusters generated in the sequencing machine, these methods provide co-methylation patterns on individual molecules within the limits of the length of the reads (up to 600 base pairs in paired-end modus on the MiSeq). With a current output of ~50 M reads for the MiSeq, yielding between 3.8 and 15 GB of sequence depending on the used sequencing kit, several tens to hundreds of target regions can be analyzed simultaneously depending on the desired coverage and number of samples analyzed in parallel. Their short run time, relatively low running costs, and wide availability make them a valuable alternative for targeted DNA methylation analysis. While initial approaches have been devised for the analysis on the pyrosequencing-based 454 platform (Taylor et al. 2007; Gries et al. 2013), which has been shown to yield accurate and quantitative results, and despite its nearly unrivaled sequencing length, this sequencing platform has not been able to cope with the improvements of other platforms and is supposed to be phased out in the near future.

In general, PCR amplification products are prepared from bisulfite-treated DNA using a two-round amplification protocol with a first pair of target region-specific primers that contain tag sequences to label the created amplicons with sequences

compatible for subsequent PCR amplification with the full-length Illumina p5 and p7 adaptor sequences. Molecular barcodes and full-length adaptor sequences are added in a second round of amplification after pooling all amplicons from a sample. If sample quantity is limited, the first amplification can also be performed as multiplex PCR, but this requires some more optimization for the multiplex setup (Korbie et al. 2015). Another strategy consists of using conventional amplification primer, and the molecular barcode and adaptors complementary to the sequences immobilized on the flow cell are subsequently added in a standard library preparation protocol after pooling of the PCR products from the first round of amplification (Jenkins et al. 2014).

An alternative protocol, termed bisulfite amplicon sequencing (BSAS), makes use of the hyperactive Tn5 transposase (Nextera/Illumina) for random insertion of the sequencing primers in the amplification product after a PCR-based targeted amplification using conventional primers (Masser et al. 2013). Amplification of multiple products in parallel can also be performed using microfluidic tools such as the Fluidigm access array, amplifying simultaneously 48 target regions in 48 samples starting from as little as 50 ng of DNA (Paliwal et al. 2013). The access array uses a two-step amplification procedure where universal forward and reverse adaptor sequences are added to the 5′-ends of the gene-specific amplification primers. Sample barcodes and platform-specific sequencing primers are added in a second round of amplification. However, this approach is unsuitable for DNA extracted from FFPE samples (Korbie et al. 2015). Bisulfite patch PCR is another approach enabling multiplex amplification by using a restriction enzyme digest to anneal exonuclease-resistant patch oligonucleotides and universal primers complementary to the created overhang, while unselected fragments are eliminated by an exonuclease digestion (Varley and Mitra 2010). Following bisulfite conversion, fragments are amplified using sequencing platform-specific universal primers. Bisulfite patch PCR has been shown to analyze up to 94 simultaneously amplified sequences with little off-target sequences; however, the requirement for a specific restriction site imposes serious limitations on the targets that can be analyzed in parallel. Nonetheless, multiple restriction enzymes can be used to select different sets of target sequences (Varley and Mitra 2010).

As for the genome-wide approaches for DNA methylation analysis, post-bisulfite adaptor tagging (PBAT) can also be applied to amplicon bisulfite sequencing approaches, allowing to start from ~100× less input DNA compared to the standard library preparation protocol and to reduce the number of amplification cycles (Miura and Ito 2015). In general, bisulfite sequencing using next-generation sequencers with their digital readout enables a more accurate quantification of DNA methylation levels, as they show a reduced error of quantitation and lower standard deviations compared to the conventional (analog) sequencing approaches (5 vs. 5–20 %) (Masser et al. 2013). Sequencing depth of ~1000 × is sufficient for a precise measurement of the DNA methylation levels, and increasing sequencing depth does not improve the accuracy further (Masser et al. 2013). However, accuracy is already very high (>99 %) if only regions with a reasonable amount of coverage (i.e., > 50 ×) are used for base calling and quantitative determination. Up to 96 samples with

different regions of interest can be currently analyzed in parallel using conventional multiplexing strategies such as (dual) indexing. The recent shift from several vendors from six base indexes to eight base indexes will further increase multiplexing capabilities. Of note, the MiSeq does already exist in a version certified for diagnostics, which potentially allows the analysis of DNA methylation-based biomarkers in a clinical setting. Specialized bioinformatic pipelines coming with the instruments or freely available pipelines, such as Bismark (Krueger and Andrews 2011) or BiQ Analyzer (Lutsik et al. 2011), enable convenient and standardized analysis of the sequencing results including the demultiplexation of individual samples, alignment to target regions, and estimation of the DNA methylation degree, allowing the analysis of DNA methylation in target regions without great bioinformatic expertise.

Ion Torrent's PGM sequencer has also been used for locus-specific DNA methylation analysis (Nones et al. 2014), and more recently this technology has been used to sequence the methylation patterns of candidate genes in circulating cell-free DNA (Vaca-Paniagua et al. 2015).

7.2 Pyrosequencing

Pyrosequencing® (Ronaghi et al. 1998; Harrington et al. 2013) is a quantitative real-time sequencing method that is frequently used for genotyping (Langaee and Ronaghi 2005) as well as for the analysis of DNA methylation patterns and allows for the accurate measurement of methylation levels in a sequence of up to 100 bp (Tost and Gut 2007; Tost et al. 2003a; Dupont et al. 2004). Pyrosequencing is based on the presence or absence of the incorporation of a nucleotide during primer extension (Ronaghi et al. 1998; Ronaghi 2001). In contrast to Sanger sequencing, which relies on the random incorporation of fluorescent ddNTPs during primer extension steps, only one specific nucleotide is present at any time in the pyrosequencing reaction. The pyrophosphate (PPi) released following nucleotide incorporation is used as a substrate in combination with adenosine 5′ phosphosulfate (APS) by an ATP sulfurylase to produce ATP (Ahmadian et al. 2006). The latter is in turn used by luciferase to oxidize luciferin into oxyluciferin resulting in a light emission that is stoichiometrically proportional to the amount of incorporated nucleotide (Ahmadian et al. 2006).

For DNA methylation analysis, a region of interest is amplified after bisulfite conversion with a standard PCR with one of the two primers being biotinylated. The biotinylated strand is captured on streptavidin-covered beads, the complementary strand is denatured and washed away, and a sequencing primer is annealed to the now single-stranded template before starting the pyrosequencing reaction. The methylation level is determined as the ratio of the signal corresponding to the incorporation of the nucleotides at a cytosine in a CpG dinucleotide corresponding to the methylated and unmethylated bases (i.e., C and T or G and A when using a reverse primer). Pyrosequencing signals report thus the average of all the molecules present in the reaction after amplification of the bisulfite-treated DNA. The limit of detection of pyrosequencing has been evaluated around 5 % for the minor allele, which is

far more sensitive than Sanger sequencing (Ogino et al. 2005). Additional advantages include the possibility to use cytosines outside CpG dinucleotides as internal control for bisulfite conversion efficiency, short time to results, and data analysis not requiring sophisticated bioinformatic expertise and resources. While the read length of the pyrosequencing run is restricted to 100–120 base pairs, longer amplification products can be analyzed by serial pyrosequencing, stripping of the DNA strand synthesized during the pyrosequencing reaction, and annealing of a new sequencing primer (Tost et al. 2006). However, the low temperature (28 °C) at which pyrosequencing is performed due to the thermal instability of some enzymatic components limits the length of the sequence that can be analyzed to ~ 300 base pairs, as longer amplification products with the low sequence complexity of bisulfite-treated DNA tend to form secondary structures that increase background signals or impede binding of the sequencing primer. The minimal amount of input DNA required for pyrosequencing-based DNA methylation analysis allowing highly accurate quantification is ~10 ng (Dupont et al. 2004). Lower amounts might still yield useful information, but reproducibility should be verified using replicate measurements when possible. Furthermore, for absolute quantification, the use of calibration curves mixing methylated and unmethylated DNA methylation standards is required. Pyrosequencing is also frequently used for the analysis of DNA from FFPE tissues (Newton et al. 2014); however, amplification products should be kept as short as possible due to the extensive degradation of the DNA during fixation and bisulfite conversion. Due to its quantitative accuracy, simplicity, and short time to results, pyrosequencing has become one of the most widely used technologies for locus-specific DNA methylation and can be considered as the current “gold” standard if loci of interest are known. Following an oxidative bisulfite conversion, pyrosequencing can also be used to determine accurately the level of hydroxymethylation in specific regions of interest (Stewart et al. 2015).

7.3 MALDI Mass Spectrometry

Mass spectrometry (MS) provides an attractive solution for nucleic acid analysis in general and DNA methylation analysis in particular, as it enables direct, rapid, and quantitative detection of DNA products measuring the molecular weight, an intrinsic physical property of each molecule, rather than relying on an indirect readout, such as a fluorescent tag. Liquid chromatography MS/MS is one of the most accurate methods to precisely quantify the global level of CpG methylation and its oxidative derivatives in samples of clinical interest (Godderis et al. 2015; Berdasco et al. 2009), while *m*atrix-*a*ssisted *l*aser *d*esorption/*i*onization *t*ime-*o*f-*f*light *m*ass *s*pectrometry (MALDI-(TOF)-MS; Karas and Hillenkamp 1988) has been the most widely used instrumental platform for the analysis of DNA methylation patterns in specific regions of interest. In this technology, the matrix, usually a low molecular weight organic acid with a strong absorption at the laser excitation wavelength, containing the analyte molecules is desorbed with a short laser pulse. The ionized nucleic acid molecules are extracted with an electric field and separated by their

masses over charge ratio in the time-of-flight to a detector. The resolution of the current generation of MALDI mass spectrometers allows an easy distinction of nucleobase substitutions in the mass range of 1000–7000 Da, which corresponds to DNA sizes of 3–25 nucleobases. The methylation status is deduced from the proportional surface area of the peaks differing by 16 Da, corresponding to the difference between formerly methylated and unmethylated cytosine nucleotides after bisulfite treatment. The dynamic range of detection of MALDI-MS is between two and three orders of magnitude, yielding highly linear responses in titration experiments and low deviations (2–3 %) from the expected values (Ross et al. 2000). Although amenable to very high-throughput through a large degree of automation and highly parallel analyses, MALDI-MS does not permit genome-wide analyses. However, due to their multiplexing capabilities, the quantitative readout of the relative abundance of products, and their simple and reliable procedure, the MALDI mass spectrometry-based assays are valuable tools for the identification and validation of methylation variable positions in a gene-targeted approach (Ragoussis et al. 2006). Therefore, MALDI-MS-based methods with their single-base resolution position themselves at the crucial follow-up stages for biomarker validation and large cohort analysis rather than biomarker discovery, as well as for large-scale investigations of candidate genes. Similar throughput for the analysis of specific CpG positions is not feasible with other available technologies for the analysis of single CpG positions.

The quantitative high-resolution scanning used in fragmentation-based approaches such as the EpiTYPER assay (Agena Biosciences, formerly Sequenom Biosciences) provides a quantitative readout for individual CpG sites in a target region of up to 600 base pairs with high accuracy (Ehrich et al. 2005). It allows to define the boundaries in which differential DNA methylation patterns can be detected and to identify specific CpGs that have the greatest diagnostic potential. Starting from ~ 1 μg of bisulfite-treated DNA per sample, the region of interest is amplified using a reverse primer with an added T7-promotor sequence, which is subsequently used for in vitro transcription to generate single-stranded RNA, which is significantly more stable than DNA in MALDI analysis. The RNA transcript is then digested with an uracil-specific enzyme to create short DNA fragments of a few nucleobases to adapt the analyte size to the optimal detection window of the instrument. Then, the fragments are purified to remove counterions interfering with the MS analysis. The RNA fragments are loaded on a SpectroCHIP Array, a holder with hydrophilic anchors preloaded with a solution of the matrix, using a piezo-pipetting device, and the matrix-embedded RNA fragments are subject to analysis on the MassARRAY Analyzer. Compared to other techniques that are able to achieve quantitative DNA methylation data of consecutive CpGs in a region of interest, its quantitative resolution of ~5 % and a similar limit of detection for the minor methylated allele fraction are only rivaled by pyrosequencing (Tost and Gut 2007) and targeted bisulfite sequencing approaches using NGS. In comparative studies, results obtained with the EpiTYPER do correlate well with those obtained by other quantitative methods such as pyrosequencing and bisulfite sequencing, but differ substantially from methylation-specific PCR (MSP)-based approaches (Alnaes et al. 2015; Claus et al. 2012). Systematic evaluation of each step in the

workflow showed that most of the variability of the experiment was induced by either the bisulfite treatment or the subsequent PCR amplification, while the cleavage and the mass spectrometric analyses contributed much less to the variability in the observed quantitative measurements (Ehrich et al. 2007; Coolen et al. 2007). In more than 90 % of the tested samples, a methylation difference of 10 % was successfully detected with the mass spectrometric assay (Coolen et al. 2007). The procedure is amenable to the analysis of DNA extracted from fresh frozen but also from FFPE tissues (Radpour et al. 2009). The EpiTYPER is one of the most widely used methods for the analysis of gene-specific DNA methylation patterns, and it has been applied to the large-scale analysis of DNA methylation patterns in cancer (Radpour et al. 2009). Due to its high-throughput capacities measuring 96–384 PCR products in parallel, it is also one of the most widely used methods for the validation of DNA methylation variation with specific phenotypes identified in epigenome-wide association studies (Zeilinger et al. 2013; Tobi et al. 2014; Zhang et al. 2014).

Once the methylation pattern has been characterized in detail, primer extension methods such as the commercial iPLEX assay (Ragoussis et al. 2006) or the GOOD assay (Tost et al. 2003b) can be employed to specifically target only those CpG sites with functional relevance or diagnostic potential in a specific biological context. The iPLEX offers routinely a degree of multiplexing in the low two-digit range, but 27-plexes have been reported (Ragoussis et al. 2006). The key to this assay lies in the combination of the primer design for the upstream PCR and primer extension assay combined with the selection of terminating dideoxynucleotides. This epigenotyping method enables the multiplexed analysis of multiple CpG sites from different promoter regions making full use of the strength of the mass spectrometer in automation and throughput. Variation of primer extension assays using competitive primer extension with oligonucleotide standards has also been used for the absolute quantification of fetal DNA in maternal plasma with high analytical sensitivity and specificity amplifying specifically hypermethylated fetal DNA resistant to a prior methylation-specific restriction digest (Nygren et al. 2010).

7.4 Methylation-Specific PCR and Its Quantitative Variations

Methylation-specific PCR (MSP) and its quantitative real-time variations allow detecting methylated molecules in the presence of an excess of normal (and usually unmethylated) DNA (Herman et al. 1996). MSP allows the amplification of virtually any CpG site after bisulfite treatment with three pairs of primers for amplification which are complementary to the former methylated, the former unmethylated sequences, or to genomic, unconverted DNA, respectively (Herman et al. 1996). The latter serves as a control for complete bisulfite conversion, but is in practice rarely included in the experimental design. Primers need to hybridize to sequences with at least two methylation variable positions (CpGs) to obtain the necessary specificity for selective amplification. The presence or absence of an amplification product analyzed on a conventional agarose gel reveals the methylation status of the

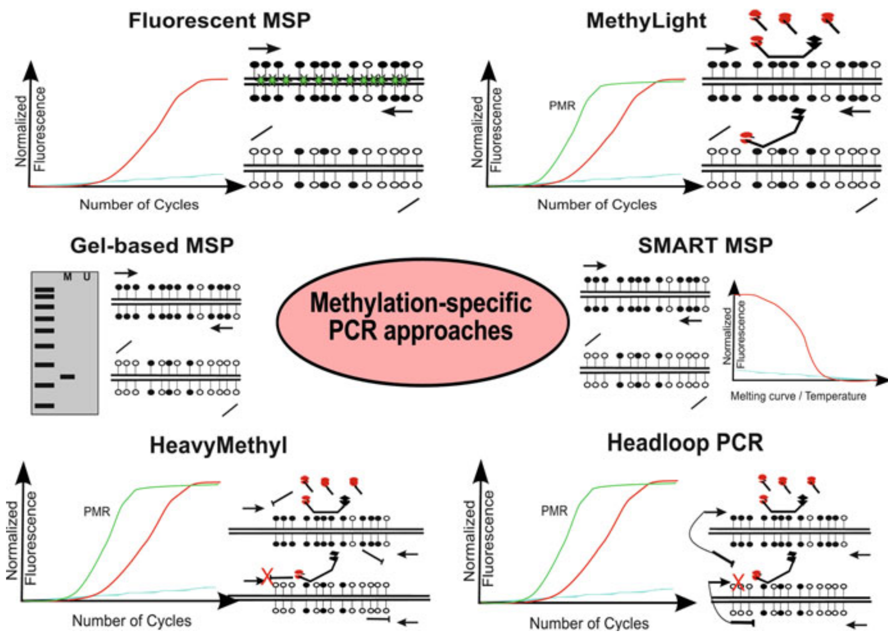


Fig. 2 Methylation-specific PCR methods for the sensitive detection of DNA methylation. For simplification only amplification with a primer complementary to a completely methylated allele is shown. *PMR* percentage of methylated allele, a completely methylated DNA standard that is used for the calculation of the percentage of methylation contained within a sample. All methods use sodium bisulfite treatment prior to PCR amplification. CpGs are depicted as lollipops, former unmethylated CpGs are shown as empty lollipops, while filled ones correspond to former methylated CpGs. Although methylation is retained as a sequence difference after bisulfite treatment, lollipops are shown for easier differentiation of the alleles. Details of the different techniques are described in the text

CpGs underlying the amplification primers (Fig. 2). MSP has been the most widely used technology for DNA methylation analysis, as it does not require any expensive instrumentation and a large number of samples can be rapidly assessed. The main advantage of MSP is the high sensitivity, which enables the detection of one allele in the presence of a 1000-fold excess of the other (Herman et al. 1996). However, MSP does not provide resolution at the individual nucleotide level, and heterogeneous methylation patterns at the primer binding sites can induce failure of amplification (Alnaes et al. 2015; Vinarskaja et al. 2012). Furthermore, the biased amplification leads to a more qualitative than quantitative result, making it difficult to distinguish different degrees of methylation at the target sites. It has been shown in comparative studies that MSP overestimates DNA methylation levels, and due to a significant number of false-positive (and negative) results, as well as the dichotomized result of MSP, associations between aberrant DNA methylation and clinical parameters are less easily identified in MSP data compared to quantitative DNA methylation technologies (Claus et al. 2012).

Quantification can be improved and false-positive results reduced by analyzing the MSP product after amplification in a real-time thermocycler by high-resolution melting analysis, an approach which has been termed SMART-MSP (Kristensen et al. 2008, 2009) for *s*ensitive *m*elting *a*nalysis after *r*eal-time methylation-specific PCR (Fig. 2). This approach makes use of a high-resolution melting analysis, which has been described in detail above and might be an alternative to the below-described real-time approaches, if no probe can be designed or if the DNA methylation patterns of the amplification product are expected to be heterogeneous, complicating the prediction of the annealing behavior of the probe. Furthermore, SMART-MSP might enable detection of DNA methylation patterns in samples difficult to amplify or detect low levels of methylation.

Real-time PCR-based methods for DNA methylation analysis, such as MethyLight (Fig. 2), use the same principle as the TaqMan® assay (Holland et al. 1991). In addition to the two amplification primers, a probe, which is dually labeled with a fluorescent reporter and a quencher dye, hybridizes to a target sequence in the amplified region and improves sensitivity, as well as specificity compared to conventional MSP. The simple one-step procedure makes real-time methylation-specific PCRs rapid high-throughput assays for quantitative DNA methylation analysis that are robust and quite resistant to carryover contamination. These approaches do not only provide information whether molecules with a certain methylation pattern are present in the sample – like conventional MSP – but also report on the fraction of them. Discrimination between methylated and unmethylated alleles can be achieved at different levels of the primers and/or the hybridization probe (Eads et al. 2000). Although in principle primers and probes could be designed for different combinations of methylated and unmethylated alleles, the most widely used approaches such as MethyLight use primers and probes that are specific for the same methylation patterns, mostly completely methylated molecules. Heterogeneous methylation patterns that display large variations between consecutive CpGs complementary to the primers or the probe will, however, also lead to a failure of the assay or biased quantitative results (Alnaes et al. 2015; Mikeska et al. 2010).

MethyLight can detect a single hypermethylated allele against a background of 10,000 unmethylated alleles (Eads et al. 2000; Lo et al. 1999). Absolute quantification of the number of molecules corresponding to the investigated pattern of methylation is achieved by measuring the ratio between the gene of interest and a reference gene, for example, the *β-actin* (*ACTB*) gene. MethyLight yields highly precise and reproducible results with an average variation of ~0.8 %, with slightly larger variations induced by different bisulfite treatments (Ogino et al. 2006). About 20-fold increased sensitivity (limit of detection and limit of quantification) can be obtained if MethyLight is not performed by conventional real-time but by digital droplet PCR (ddPCR) (Wiencke et al. 2014; Yu et al. 2015b). Furthermore, with ddPCR the quantitative accuracy is increased about sevenfold for some assays (Wiencke et al. 2014). The concept of using MethyLight on single molecules was devised some time ago using limiting dilution to perform the individual MethyLight reactions (Weisenberger et al. 2008). The use of commercially available ddPCR machines has significantly streamlined and simplified the assay procedure and the

readout. Furthermore, absolute quantification of methylated alleles can be achieved by simply counting the positive droplets in ddPCR and the use of Poisson statistics without the need for a standard curve. Hence, amplification efficiency is less of a concern compared to the conventional MethyLight. This method might therefore be better suited if very few methylated alleles are expected to be present in a clinical sample. As ddPCR enters clinical laboratories for various applications including mutation detection and quantification, these instruments will probably be widely available. Additionally, its simplified technical use and the described advantages make the technology one of the promising approaches for DNA methylation analysis entering the field of personalized medicine. Nonetheless, these assays are significantly more expensive than normal MethyLight assays, and the throughput with current ddPCR machines will be much reduced. The addition of a second probe marked with a different fluorescent dye to the qPCR-based MethyLight assay allows for the simultaneous detection of unconverted sequences that might co-amplify with the bisulfite-converted molecules avoiding potential false-positive results (ConLight) (Rand et al. 2002). Quantitative analysis of methylated alleles (QAMA) uses a TaqMan probe conjugated to a minor groove binder for discrimination at single-base level by forming hyperstabilized duplexes with complementary DNAs (Zeschnigk et al. 2004). Methylated and unmethylated alleles are simultaneously quantified using two probes modified with two different fluorophores. Thus, amplification of the bisulfite-treated DNA can be carried out with primers amplifying both the formerly methylated and the unmethylated alleles, and differentiation of the methylation status of alleles is achieved only at the probe level.

*M*ethylation-*s*pecific *fl*uorescent *a*mplicon *g*eneration (MS-FLAG) is a real-time methylation-specific PCR. Instead of relying on a probe, the fluorescent label and quencher are incorporated in the 5′ tag of the methylation-dependent amplification primers (Bonanno et al. 2007). The synthesis of the complementary strand leads to a double-stranded DNA and the creation of a recognition site for the thermostable restriction enzyme PspGI, which cleaves part of the tag off and separates the quencher from the fluorophore, thus leading to a measurable increase in fluorescence. While this method is perhaps more cost-efficient, as no probe is required, it will have lower sensitivity compared to MethyLight and has rarely been reported for DNA methylation analysis.

HeavyMethyl further increases sensitivity and specificity of real-time PCR-based assays for the analysis of DNA methylation using methylation-dependent blocking oligonucleotides (Cottrell et al. 2004). In contrast to MethyLight, amplification primers are not specific for a certain methylation pattern but positioned in sequence stretches containing no CpG positions (Fig. 2). Only the fluorescent probe is specific usually to a consistently hypermethylated sequences. The increased specificity and sensitivity are achieved through a second pair of non-extendable (3′ phosphorylated) oligonucleotides that hybridize specifically to a methylation pattern opposite to the investigated one, usually the unmethylated sequence. The annealing sites of these oligonucleotides overlap with the target sequences for PCR amplification and thereby efficiently block any amplification of the bisulfite sequence corresponding to the undesired methylation pattern. HeavyMethyl was able to specifically detect

25 pg of in vitro methylated DNA in the background of 400 ng of unmethylated DNA (relative sensitivity up to 1:8000). Similarly to MethyLight, a reference standard is used to identify samples with negligible amounts of methylation. The use of four to five different oligonucleotides contributes significantly to the cost of the assay, and the design might be more complex compared to the conventional MSP or MethyLight. However, HeavyMethyl shows the necessary sensitivity and specificity required for clinical applications, and it is the underlying principle of the commercial DNA methylation-based diagnostic tests for the (early) detection of colorectal cancer (Epi proColon, Epigenomics AG) (Church et al. 2014), targeting DNA methylation in the second intron of the *Septin9* gene, or lung cancer (Epi proLung, Epigenomics AG) (Ilse et al. 2014), targeting *SHOX2*.

An alternative approach uses amplification primers that carry a 5′-tail sequence complementary to a sequence that is present in the amplicon corresponding to a specific methylation pattern (Headloop PCR) (Rand et al. 2005). After incorporation of the primer in the synthesized PCR product, the tail folds back onto the template creating a secondary structure refractory to amplification (Fig. 2). For example, if the tail is complementary to an unmethylated sequence, only methylated molecules are amplified. The amplification is monitored in real time with *SYBR Green* or TaqMan probes. The sensitivity of the approach is similar to the others described above detecting a methylated allele in the presence of a 4000-fold excess of unmethylated ones.

Instead of using primers complementary to a specific methylation pattern, the *co*amplification at *l*ower *d*enaturation temperature (COLD)-PCR-based approach makes use of the different GC content of the unmethylated and methylated molecules after bisulfite conversion. By lowering the denaturation temperature of the PCR, only the unmethylated molecules, which have a lower GC content, will be efficiently amplified, leading to an enrichment of the unmethylated sequences in a context of mainly methylated molecules (Castellanos-Rizaldos et al. 2014).

As the amplicons of the different MSP variants are small (~100 bp), these methods usually work well with DNA of lower quality, such as DNA extracted from FFPE samples (Herman et al. 1996). No special equipment is required for the conventional methylation-specific PCR, and real-time PCR machines are available at most research institutions for the quantitative analysis of gene expression. The similarity of the approaches described below to the real-time expression analysis also facilitates implementation of the technology, execution of the experiments, and interpretation of the results in laboratories not yet very familiar with DNA methylation analysis. The design of assays and optimization of amplification are probably the most important steps to ensure specific amplification of the desired locus. Sensitivity and specificity vary largely between assays depending on the primers (and probes in case of techniques like MethyLight) and conditions. Quantitative MSP assays allow high-throughput screening of a large number of clinical specimens in a single PCR step without complicated downstream analysis. Multiplexing with methylation-independent controls normalizes for DNA input, and parallel processing of calibration standards will allow assessment of run-to-run variability. However, these assays do not provide information on the DNA methylation level of individual CpGs, and

heterogeneous DNA methylation patterns can lead to a high rate of false-positive results (Mikeska et al. 2010; Claus et al. 2012; Alnaes et al. 2015).

8 DNA Methylation Analysis of Cell-Free Circulating DNA

DNA methylation has received a lot of attention in recent years, because of its potential as a stable and amplifiable biomarker for early diagnosis, prognosis, or response to treatment in various cancers and potentially other complex diseases (How Kit et al. 2012). Biomarkers capable of distinguishing a disease state from healthy individuals must be specific, sensitive, and detectable in specimens obtained through minimally invasive procedures to be clinically applicable. Disease-specific DNA molecules can be found in various body fluids, such as urine or sputum, or as circulating cell-free (ccf) DNA molecules that can be isolated from the serum/plasma of cancer patients (Schwarzenbach et al. 2011; Heitzer et al. 2015; Diaz and Bardelli 2014), individuals with autoimmune diseases (Chan et al. 2014), as well as individuals with many other complex diseases and physiological conditions under which cells undergo apoptosis and shed DNA molecules into the bloodstream (Lehmann-Werman et al. 2016).

Few of the so far described methods have the potential to be used for the detection of DNA methylation in body fluids as the target DNA molecules are present at only very low concentrations among an excess of DNA from healthy cells. Furthermore, only part of the molecules will carry the DNA methylation patterns of interest. Therefore, high analytical sensitivity of the method is crucial to detect the low levels of tumor-derived aberrantly methylated DNA molecules that are present in these specimens. The analytical specificity, i.e., the frequency of false positives obtained with the method, is also primordial for the use of a method in the clinics. It is therefore important to verify that the target regions are not methylated even at low levels in leukocyte DNA. Methylation-specific PCR (Hoon et al. 2004) and particularly methylation-specific real-time PCR-based methods such as MethyLight (Campan et al. 2011; Begum et al. 2011), HeavyMethyl (Church et al. 2014), as well as methylation-specific high-resolution melting analysis (MS-HRM) (Yang et al. 2015) and MRE-qPCR (Wielscher et al. 2015) have proven suited for the detection of very low levels of aberrant methylation in circulating DNA (see also Table 2). The commercial Epi *pro*Colon test, which has been approved by the Chinese FDA in July 2015 and the US FDA in April 2016, analyzes methylation in the *SEPT9* gene in cell-free circulating DNA for the population-wide screening for colorectal cancer (Warren et al. 2011; Church et al. 2014), using the HeavyMethyl technology. Gene-specific assays analyzing DNA methylation changes in cell-free circulating DNA have recently been reviewed in detail (Warton and Samimi 2015).

Next-generation sequencing approaches are becoming more and more used to identify and monitor the presence of mutations in the cell-free DNA isolated from plasma (Newman et al. 2014; Crowley et al. 2013; Heitzer et al. 2015). Sequencing approaches of bisulfite-treated DNA isolated from plasma or serum are complicated by the fact that the bisulfite treatment will further degrade the DNA fragments

reducing the amount of amplifiable DNA, but a number of successful examples have now been documented (Lun et al. 2013; Chan et al. 2014). Most of the approaches for the whole-genome bisulfite sequencing described in this chapter are compatible with the DNA amount that can be obtained from a few mL of plasma, and the only adaptation that is required is the omission of the fragmentation step, as the cell-free DNA is already in a convenient size range. Methylated DNA can also be enriched using a methyl-binding protein as in the MBD-seq approaches prior to NGS, but this requires relatively large volumes of plasma to obtain the required starting amount of 50 ng of cell-free DNA (Warton et al. 2014). DNA methylation is an attractive marker for the analysis of cell-free DNA as DNA methylation changes are widespread in many diseases, particularly in cancer. It is, therefore, more likely that sequencing of the plasma DNA will detect these changes when compared with the detection of more specific focal changes such as specific copy number alteration or a single-nucleotide mutation. Hypomethylation of cell-free DNA isolated from plasma, as assessed by sequencing of bisulfite-treated DNA, yielded a diagnostic sensitivity of 74 % and a specificity of 94 % for nonmetastatic hepatocellular cancer cases (Chan et al. 2013a), being superior to the analysis of repetitive elements such as LINE1 retrotransposons in cell-free DNA (Tangkijvanich et al. 2007). Differences were assessed by binning of the methylation densities (reads) in 1 MB intervals and comparing methylation levels between patients and controls. As little as 10 M sequencing reads per sample were sufficient to obtain a similar sensitivity and specificity. Copy number alterations can be accurately deduced from low-coverage sequencing of cell-free circulating DNA (Chan et al. 2013b; Heitzer et al. 2013). As these copy number changes are retained in the bisulfite-treated DNA, bisulfite sequencing will inform on both the methylation status and the copy number status without additional cost for the copy number analysis (Chan et al. 2013a).

While most of the work has so far been performed in cancer, recent data show the potential of bisulfite sequencing for other complex diseases as well. Examples include autoimmune diseases, where the genome-wide hypomethylation observed normally in blood cell population was also detected in the sequenced cell-free DNA (Chan et al. 2014) or a number of other complex diseases where DNA methylation patterns specific to the diseased organ could be detected in cell-free circulating DNA. Examples of methylation changes detected in ccf DNA in recent works are (1) methylation changes at the insulin or amylin promoter specific for pancreatic β-cells in type 1 diabetes patients, (2) changes at *MBP3* and *WM1* specific for oligodendrocytes in relapsing multiple sclerosis patients, (3) changes in methylation of the brain-specific CpG CG09787504 locus in brain cells of patients after traumatic or ischemic brain damage, (4) changes in methylation at PPARγ that reflect the extent of liver fibrosis in patients with nonalcoholic fatty liver disease, and (5) changes at *CUX2* and *REG1A* in exocrine pancreas cells of patients with pancreatic cancer or pancreatitis (Lehmann-Werman et al. 2016; Olsen et al. 2016; Hardy et al. 2016).

A more specialized area for the analysis of DNA methylation patterns in cell-free circulating DNA is prenatal diagnosis. Analysis of differences in the DNA methylation patterns between the maternal and fetal circulating DNA molecules has been

proposed as an alternative strategy to the analysis of DNA sequence-based variations (Chu et al. 2009; Papageorgiou et al. 2009; Nygren et al. 2010). Circulating cell-free fetal DNA is shed through apoptosis from the placenta and can be isolated from maternal plasma or serum from 5 weeks gestation. A major analytical problem for the analysis of the circulating cell-free fetal DNA is the high risk of false-negative results due to failure to extract or detect sufficient material and/or due to large individual variability in the total amount of cell-free DNA and the contribution of the fetal component to this total amount. Several studies have identified a number of genomic regions that are differentially methylated between the fetal DNA derived from the placental tissue and the maternal peripheral blood mononuclear cells (Chim et al. 2008; Papageorgiou et al. 2009). This epigenetic information can be used for diagnostic purposes by isolating the fetal DNA using methylation-specific restriction enzymes (Chim et al. 2005). Thus, if the sequence is only methylated in the fetal DNA, only the fetal DNA will be amplifiable after the digestion addressing two major challenges associated with the analysis of circulating cell-free fetal DNA.

MALDI mass spectrometry has been used to analyze placenta-specific DNA methylation profiles of genes located on different chromosomes to accurately determine the fetal portion of the circulating cell-free DNA isolated from maternal plasma and presented the first universal biomarker for the fetal DNA quantification (Nygren et al. 2010). Target regions are co-amplified in the presence of known quantities of synthetic templates differing by a single nucleotide from the target regions, enabling the accurate quantification of the total number of fetal copies as well as the fraction of the fetal DNA in the maternal plasma with high specificity and sensitivity of 100 % and 99 %, respectively (Huang et al. 2006; Tsui et al. 2005).

Bisulfite sequencing of the fetal methylome has also been demonstrated using genetic differences to separate maternal- and fetal-derived molecules after sequencing. It allowed the analysis of ~ 100,000 loci and covered ~ 200,000 CpGs showing a strong resemblance of the fetal and placental methylomes (Lun et al. 2013). An alternative approach made use of the specific placental differentially methylated regions (Papageorgiou et al. 2009; Chim et al. 2008), which had an opposite methylation patterns in blood cells (Lun et al. 2013). In addition to the above-described approaches using Illumina sequencers, promoter regions of selected candidate genes have also been sequenced on the Ion Torrent PGM sequencer starting from cell-free circulating DNA (Vaca-Paniagua et al. 2015).

9 Single-Cell DNA Methylation Analysis

Due to the large divergence of DNA methylomes between cell types and the cellular heterogeneity of tissues, the recent advances in single-cell transcriptomics (Saliba et al. 2014; Jaitin et al. 2014; Klein et al. 2015) and the insights gained from these studies have raised a lot of interest for single-cell DNA methylation analysis.

The first prerequisite for working with single cells is to capture them. This is efficiently achieved with the state-of-the-art cell isolation techniques. Sample preparation techniques for DNA methylation have been combined with multiplex assays

using the SCRAM assay, which applies methylation-specific restriction enzymes after single-cell isolation and lysis, and locus-specific qPCR on a Fluidigm Biomark system, which allows the interrogation of 24 genomic positions in 48 single cells in one experiment (Lorthongpanich et al. 2013; Cheow et al. 2015). By using methylation-specific restriction enzymes, this approach avoids the degradation of the DNA by the bisulfite conversion, which can impede subsequent amplification if only little material is available, as in the case of single cells. A single-cell DNA methylation analysis method that combines bisulfite conversion with Sanger sequencing and Agena's EpiTYPER also allows cost-effective analysis of a larger number of single cells (Gravina et al. 2015).

Recently, techniques combining single-cell analysis with the second-generation sequencing have become a major field of development. Apart from the technical challenges, the main issue with single-cell analysis is the number of cells that needs to be analyzed to obtain a representative picture of the biology and the amount of sequencing that can be afforded per cell. A balance needs to be struck between these two parameters. Single-cell whole-genome bisulfite sequencing has been shown using the PBAT approach and a whole-genome pre-amplification step (Smallwood et al. 2014). However, this approach has limited genomic coverage per cell, as it only recovers 48.4 % of all CpG positions for each cell analyzed. This means that many more cells need to be included to capture subpopulations, and large oversampling is necessary. On the other hand, it allows capturing non-CpG methylation. The pre-amplification step can be omitted, as shown in μWGBS approach, but yields libraries with lower complexity and therefore lower coverage (Farlik et al. 2015). An alternative to preparing libraries for sequencing from individual cells is to generate pools of a small, defined number of cells, sequence each pool, and then use computational methods to deconvolute cell states due to the distortion in methylation detected between pools (Farlik et al. 2015). RRBS has been applied to single cells and is a good compromise in terms of genomic coverage, suitability with low input, and reduction of the target size allowing to assess the cell-to-cell heterogeneity of DNA methylation patterns, if no comprehensive coverage of the genome is required (Guo et al. 2013, 2015). It also allows for the quantitative analysis of differential DNA methylation (Wang et al. 2015a).

Complementary to the above-described single-cell technologies, DNA methylation has also been detected in nanofluidic channels at the single-molecule level using fluorescent labeling of methylated DNA with MBD1 (Cipriany et al. 2010, 2012). This method combined with fluorescence-activated cell sorting, termed SCAN for *si*ngle *c*hromatin molecule *a*nalysis at the *n*anoscale, allows also selecting molecules with distinct epigenetic patterns for further analysis using higher-resolution technologies, such as qPCR and potentially NGS (Cipriany et al. 2012). It also offers the possibility of the simultaneous analysis of DNA methylation and chromatin modifications, as the labeling of the DNA and size of the nanochannels allow to isolate DNA together with its native chromatin, as described in more detail in Sect. 15.12.

Although all the methods do provide a picture of the DNA methylation state at a given time point, they do not allow monitoring the dynamics of the DNA

methylation at the single-cell level. This can be achieved with a GFP-coupled reporter construct for locus-specific DNA methylation changes with single-cell resolution that can be inserted at specific loci using the CRISPR/Cas9 technology (Stelzer et al. 2015). The methylation state of the surrounding sequence will influence the methylation state of the reporter construct, thereby allowing tracing the methylation level of the endogenous sequences and its dynamic changes during development and disease.

10 Analysis of Hydroxymethylation

Although 5-hydoxymethylation has been known since several decades, it has recently attracted much more attention, as it constitutes an intermediate in the active DNA demethylation process and is thought to play an active role in the regulation of gene expression. Furthermore, altered patterns of hydroxymethylation have been found in different diseases, notably cancer and neurodegenerative diseases (Delatte et al. 2014; Wang et al. 2014). In general, the total levels of 5-hydroxymethylcytosine observed across genomes are approximately tenfold lower compared to 5-methylcytosine, although large variations between tissues exist (Ruzov et al. 2011a). With the surge in interest to determine the exact location and relative abundance of hydroxymethylation, several technologies have been developed permitting its genome-wide or locus-specific analysis (Table 3; Fig. 3). However, although the potential confounding of DNA hydroxymethylation and methylation by bisulfite-based methods has been recognized early on (Nestor et al. 2010; Jin et al. 2010), most studies still tend to ignore this problem for both reasons of cost, but also because of the lower prevalence of 5hmC.

A major difference between 5-methylcytosine and 5-hydroxymethylcytosine is the possibility of the latter to be modified by glycosylation (Fig. 3). 5hmC is a naturally occurring base in some bacteriophages and is in these organisms often further modified by glycosylation by glycosyltransferases as a defense mechanism against digestion by restriction endonucleases present in the host (Vrielink et al. 1994). However, a number of restriction enzymes have recently been reported that specifically recognize and cleave 5hmC-containing sequences after glycosylation, with PvuRts1I being the first enzyme identified (Borgaro and Zhu 2013; Wang et al. 2011a) (Fig. 3). These enzymes cleave at a defined distance, normally 11–13 nucleotides 3′ from the modified cytosine. Aba-seq uses the enzymatic properties of AbaSI (AbaSDFI), a member of the PvuRts1I restriction enzyme family shown to exhibit high and improved specificity for 5hmC over 5mC and C compared to PvuRts1I (Wang et al. 2011a). Starting with 2 μg of DNA, 5hmCs are glycosylated and subsequently cleaved, prior to the ligation of biotinylated adaptors, which allow the subsequent capture of sequences containing hydroxymethylated fragments. While initially only locus-specific analyses have been performed using this approach (Wang et al. 2011a), recent studies also demonstrated its applicability to genome-wide analyses using NGS starting from as little as 100 ng of input DNA (Sun et al. 2013; Gross et al. 2015). This method allows to cover ~ 58 % of all potentially

Table 3 Technologies for the genome-wide analysis of 5-hydroxymethylcytosine

Technology	The technology behind	Single-nucleotide resolution (yes/no)	Absolute quantification (yes/no)	DNA input	Required number of reads per sample in M	Coverage of CpGs (theoretical)	Comments	References
Aba-seq	Glycosylation of 5hmC, followed by restriction digest with AbaSI, ligation of biotinylated adaptors, size reduction of fragments, and library preparation	Yes, for most fragments (82 %)	No	100 ng–2 µg	200 M	58 %	Simple assay procedure, identifies 5hmC in regions with little 5hmCs, replicate measurements required for good coverage	Sun et al. (2013), Gross et al. (2015)
RRHP	Genomic DNA is digested with Msp1, 5hmC is glycosylated, and DNA is digested again with Msp1, prior to size selection and sequencing	Yes, but only at restriction sites	No	500 ng	20–30 M	15 % (mainly in CpG-rich regions)	Reproducible data, but limited to Msp1 sites, limited coverage in gene regulatory regions with low CpG density	Petterson et al. (2014)
HELP-GT	The sample is divided in three aliquots and digested with either HpaII, MspI, or Msp1 after glycosylation, prior to library preparation and sequencing	Yes, but only at restriction sites	Yes	300 ng	30 M	10–15 % (mainly in CpG-rich regions)	Parallel analysis of unmethylated, methylated, and hydroxymethylated cytosines allows for quantification, but limited to Msp1 sites, limited coverage in gene regulatory regions with low CpG density	Bhattacharyya et al. (2013)

5hmC-MeDIP	Immunoprecipitation with an antibody specific for 5-hydroxymethylcytosine	No	No	1–4 µg	20–40 M	100 %	Preferential enrichment of 5hmC in regions with high density of 5hmCs and certain sequence contexts, low (interlaboratory) reproducibility	Williams et al. (2011), Wu et al. (2011), Thomson et al. (2013)
CMS-MeDIP	Immunoprecipitation with an antibody specific for 5-cytosinemethylene-sulfonate, an intermediate product of the bisulfite conversion of 5hmC	No	No	5 µg	40 M	100 %	Increased reproducibility compared to 5hmC-MeDIP, but still biased for hydroxymethylation density	Pastor et al. (2011)
JBP1-seq	Glycosylation of 5hmC, followed by enrichment with JBP1	No	No	50 ng–1 µg	10–40 M	100 %	Lower specificity and sensitivity, high background compared to other enrichment methods, recent improvements using recombinant JBP1 proteins	Robertson et al. (2012), Cui et al. (2014)
GLIB	Glycosylation of 5hmC followed by oxidation and attachment of a biotin group used for subsequent enrichment and sequencing	Yes/No	No	1 µg	40 M	100 %	No dependency on hydroxymethylation density, complex multi-step protocol, only relative enrichment can be measured	Pastor et al. (2011)

(continued)

Table 3 (continued)

Technology	The technology behind	Single-nucleotide resolution (yes/no)	Absolute quantification (yes/no)	DNA input	Required number of reads per sample in M	Coverage of CpGs (theoretical)	Comments	References
hMeSeal	Glycosylation of 5hmC with an azide containing glucose moiety followed by attachment of a biotin group used for subsequent enrichment and sequencing	Yes/No	No	1 µg	40–80 M	100 %	Streamlined protocol, chemical reagents available as a kit, only relative enrichment can be measured	Song et al. (2011)
TAB-seq	Glycosylation of 5hmC, oxidation of all methylated cytosines to 5caC with a recombinant TET enzyme (5hmCs are protected), bisulfite treatment followed by NGS sequencing	Yes	Yes	3 µg	600 M	100 %	Quantitative high-resolution 5hmC profiles, requires large amount of sequencing to get accurate estimation of 5hmC levels at all cytosines, direct readout of 5hmC, more complex protocol compared to OxBS-seq	Yu et al. (2012)

OxBS-seq	Part of the samples is analyzed by standard (whole genome) bisulfite sequencing, while in the other part 5hmC is oxidized to formylcytosine prior to bisulfite conversion. 5hmC levels are obtained by subtracting methylation levels from both reactions	Yes	Yes	400 ng–1 µg	1200 M	100 %	Quantitative high-resolution 5hmC profiles, requires large amount of sequencing to get accurate estimation of 5hmC levels at all cytosines as WGBS is required in addition (doubled compared to TAB-seq), combination with Infinium BeadChips or RRBS is an alternative strategy, 5hmC levels are inferred, not measured, potential oxidative damage to DNA	Booth et al. (2013), Field et al. (2015)

Abbreviations: *5hmC* 5-hydoxymethylcytosine, *Aba-seq* AbaSI-coupled sequencing, *RRHP* reduced representation hydroxymethylcytosine profiling, *HELP-GT* *H*paII tiny fragment *e*nrichment by *l*igation-mediated *P*CR-*g*lycosyl *t*ransferase assay, *MeDIP* *me*thylated *D*NA *i*mmuno*p*recipitation, *CMS* cytosine methylene sulfonate, *GLIB* *gl*ucosylation, per*i*odate oxidation, *b*iotinylation, *hMeSEAL* 5hmC-selective chemical labeling, *TAB-seq* TET-assisted bisulfite sequencing, *OxBS-seq* OxBS: *ox*idative *b*isulfite *s*equencing

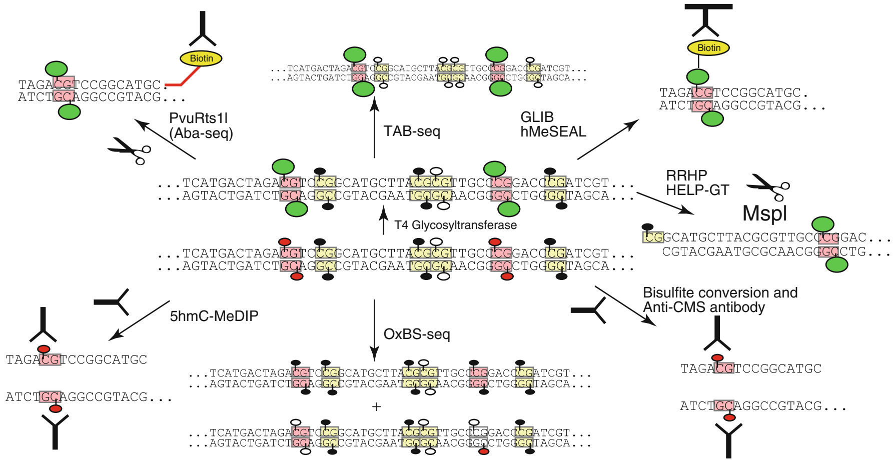


Fig. 3 Overview (simplified) of the different commonly used assays for the analysis of 5-hmC. Details are given in the text and in Table 3. CpGs are depicted as lollipops, unmethylated CpGs are shown as empty lollipops, while filled ones correspond to methylated CpGs; lollipops filled in red denote 5-hydroxymethylcytosines, while those in green depict hydroxymethylated and glycosylated cytosines

hydroxymethylated cytosines in the genome (Wang et al. 2011a). In contrast to the affinity-based enrichment methods, where the read numbers correlate with the density of hydroxymethylation, Aba-seq might be better suited for the identification of regions containing few hydroxymethylation marks, despite its limitation in coverage (Sun et al. 2013).

In addition, conventional restriction enzymes such as MspI can be used to differentiate between the two nucleosides after glycosylation of 5-hydroxymethylcytosines as the endonuclease activity of MspI is blocked by the glucosylation, but not methylation at the CpG cytosine (Kinney et al. 2011). *R*educed *r*epresentation 5-*h*ydroxymethylcytosine *p*rofiling (RRHP) makes use of the same principle (Petterson et al. 2014). Genomic DNA is digested twice with MspI; digestion of the DNA is separated by a 5hmC glucosylation step prior to the size selection and sequencing. RRHP exploits the β-glucosyltransferase to inhibit the enzymatic cleavage of the adapters ligated to a genomic library, allowing only fragments with glucosylated 5hmC residues at adapter junctions to be amplified and sequenced, thus providing a positive display of hydroxymethylation (Fig. 3). While this protocol is quite robust and yields high-quality data with relatively few reads (20–30 M), allowing a large number of samples to be simultaneously analyzed on a HiSeq instrument, the dependence on MspI restriction sites allows to analyze only ~15 % of CpG sites in the human genome. Using alternative restriction enzymes allows the analysis of cytosines in other sequence contexts including non-CpG methylation (Sun et al. 2016). The analysis of 5hmC using the HELP assay (Khulan et al. 2006) makes use of the same principle: a part of the sample is glycosylated prior to the standard MspI digestion included in the protocol and allows in comparison with the standard

MspI digest (without prior glycosylation) to determine hydroxymethylated cytosines (Bhattacharyya et al. 2013).

In contrast to the bisulfite-based methods, affinity-based enrichments such as MeDIP, MIRA, and several of the MBDs including MBD1, MBD2, and MBD4 and MeCP2 do not recognize hydroxymethylated cytosines (Jin et al. 2010). A variety of new affinity-based methods have been devised to profile specifically 5-hydroxymethylation genome-wide using either (1) antibodies against 5hmC or an intermediate product of its bisulfite conversion (5-methylenesulphonate), (2) glucosylation of 5hmC followed by biotinylation, and (3) specific pull-down of glycosylated methylated DNA with a JBP1 protein (Fig. 3).

Similar to the above-described MeDIP-seq protocols for the enrichment of methylated cytosines, antibodies raised against hydroxymethylated cytosines have been used to profile 5-hydroxymethylation genome-wide (Williams et al. 2011; Ficz et al. 2011; Wu et al. 2011). However, the antibodies were subsequently found to enrich preferentially regions with a high density of hydroxymethylated cytosines and to display large interlaboratory variations. In addition, there might be an enrichment of some sequence contexts such as tandem repeat sequences (Pastor et al. 2011; Matarese et al. 2011; Thomson et al. 2013). The treatment of 5hmC with sodium bisulfite yields 5-cytosinemethylenesulfonate (CMS) as an intermediate product, which can be used to isolate sodium bisulfite-converted 5hmC using anti-5-methylenesulfonate antibodies (Pastor et al. 2011). As an alternative strategy, the GLIB approach involves the glucosylation of 5-hydroxymethylcytosines, oxidation with periodate, and biotinylation of 5mhC (Pastor et al. 2011). A glucose moiety is added to 5hmC by a glucosyltransferase, and the vicinal hydroxyl groups are subsequently oxidized to aldehydes by treatment with sodium periodate. Afterward, biotin molecules are added to the newly formed aldehyde groups (Pastor et al. 2011). The GLIB approach allowed for the effective pull-down of >90 % of fragments containing a single 5hmC, while the anti-CMS-based enrichment was more dependent on the density of 5hmC, but still achieved lower background levels compared to the conventional 5hmC-MeDIP (Pastor et al. 2011).

In a similar approach to GLIB, a glucose moiety containing an azide group can subsequently be used in a click chemistry step to attach biotin molecules to 5hmC positions and enrich and sequence 5hmC-containing sequences (Song et al. 2011). This approach requires fewer steps and induces less DNA damage when compared to GLIB (Song et al. 2011). In direct comparisons, approaches based on chemical labeling proved to be more specific than antibody-based methods (Thomson et al. 2013). For the quantification of the global level of 5hmC, a similar approach transferring a radioactively labeled glucose moiety can be used (Szwagierczak et al. 2010). The selective chemical labeling (SLC) exonuclease protocol combines the glycosylation and azide-based biotinylation with an exonuclease digestion, where the exonuclease is blocked at the first modified 5hmC (Serandour et al. 2016). While this protocol allows the determination of hydroxymethylation in all sequence contexts at single CpG resolution, several replicates of this multi-step procedure are required to obtain a reliable and comprehensive coverage.

A last approach uses the trypanosome J binding protein 1 (JBP-1) to enrich for glycosylated hydroxymethylated cytosines (Robertson et al. 2011, 2012). However,

the protocol is more time-consuming and complicated, and the enrichment is much less efficient compared to hMeDIP or chemical labeling, leading to a high background noise and a number of false-positive enrichment peaks (Thomson et al. 2013). Recent improvements using an in vivo biotinylated JBP1 protein show increased specificity compared to the original approach and allow starting from as little as 50 ng of DNA (Cui et al. 2014). While this protocol has been commercialized, no independent evaluation with other technologies has so far been performed.

However, most of the above-described methods do not provide single-nucleotide resolution and due to the enrichment step do not allow for precise quantification of the hydroxymethylation level. The relative low abundance of 5hmC requires sensitive and preferably single-nucleotide resolution methods for its detection. Currently, two approaches are commonly used: TAB-seq and oxidative bisulfite sequencing.

The TET-assisted bisulfite sequencing (TAB-seq) protocol again uses the differential potential of hydroxymethylated and methylated cytosines to undergo glycosylation. Hydroxymethylated cytosines can be glycosylated and are thus protected from oxidation with recombinant TET enzyme, which converts methylated cytosines to carboxymethyl-cytosines, which are subsequently deaminated to uracils using a conventional bisulfite treatment (Yu et al. 2012). Thus, TAB-seq allows a positive readout of hydroxymethylation at single-nucleotide resolution, as all remaining cytosines in the bisulfite-converted sequences should correspond to hydroxymethylated cytosines (Fig. 3). Of note, TAB-seq has also been combined with Roche's SeqCap Epi CpGiant capture probes to sequence regions of interest to a much higher depth (Li et al. 2015c).

Oxidative bisulfite (OxBS-seq) sequencing is based on the selective and efficient oxidation of 5-hydroxymethylated cytosines to 5-formylcytosines, while 5-methylcytosines are resistant to the oxidation reaction (Booth et al. 2012, 2013). In a subsequent bisulfite conversion, 5-formylcytosines are deaminated to uracils, like unmodified cytosines, and only methylated cytosines appear as cytosines in the sequence readout. Levels of hydroxymethylation can therefore be deduced by subtracting methylation levels of the OxBS-seq reaction from a standard bisulfite sequence (Fig. 3). This approach has created considerable interest, as it allows adapting current workflows for whole-genome bisulfite sequencing for the detection of 5-hydroxymethylcytosine. However, it should be underlined that OxBS-seq necessitates a standard bisulfite-treated reference sequence, which requires that two whole-genome bisulfite sequencing analyses have to be performed. Furthermore, due to the low abundance of 5-hydroxymethylation in most tissues, an increased coverage might be required to reliably identify changes in hydroxymethylation, which needs to be achieved in both reactions.

As described above in the paragraph on technologies for the genome-wide analysis of DNA methylation patterns, the 450 K Illumina Infinium BeadChips are currently a very popular technology for the analysis of DNA methylation patterns in humans, especially for studies analyzing large cohorts, and two different approaches have been devised to use this platform also for the analysis of hydroxymethylation. The 450 K BeadChip also allows for a genome-wide analysis of hydroxymethylation at a fraction of the cost of the whole-genome sequencing, and bioinformatic

analysis is simplified. In the TAB-array approach, the TAB-seq principle (Yu et al. 2012) is transferred to the 450 K BeadChip (Nazor et al. 2014; Chopra et al. 2014). The fragmented DNA after glucosylation and bisulfite conversion is analyzed on the methylation BeadChips using the standard experimental procedure. As TAB-array yields a positive readout of hydroxymethylation, but hydroxymethylation is a relatively rare DNA modification, the distribution of the observed beta values is profoundly different from the one obtained by the conventional bisulfite-based analysis with a nearly complete unimodal distribution close to a beta value of zero. Therefore, separate normalizations for beta values obtained by bisulfite and TAB-array need to be performed. The 450 K BeadChips have also been combined with oxidative bisulfite sequencing using the bisulfite-converted and OxBS-converted DNA as input into the standard Illumina workflow (Stewart et al. 2015; Field et al. 2015). The overall levels of hydroxymethylation correlated well with the hydroxymethylation levels measured by liquid chromatography coupled with mass spectrometry and the results obtained by qPCR in conjunction with glycosylation and restriction enzyme digestion (Stewart et al. 2015; Field et al. 2015). Furthermore, the hydroxymethylation patterns differed by less than 10 % when compared to OxBS pyrosequencing (Stewart et al. 2015). As at individual CpG positions the level of hydroxymethylation might be much lower than the level of cytosine methylation and the BeadChip measurement do have some inherent technical variability, experiments were performed in quadruplicates to ensure the reliable detection of hydroxymethylation. The number of detectable hydroxymethylated CpG sites decreased by half when only duplicates were performed (Field et al. 2015). While too few samples have been analyzed on the 450 K BeadChip in any of the so far published studies to draw definite conclusions, the degree of correlation with other methods seemed to be slightly higher for the studies using the OxBS approach compared to the TAB-array. The increased content of intergenic and enhancer sequences on the newly devised Illumina EPIC array makes this novel version of the BeadChip even more interesting for the analysis of 5-hydroxymethylation (Moran et al. 2016). Due to the highly quantitative nature of its readout (Tost and Gut 2007), the pyrosequencing technology is also ideally suited for the analysis of gene-specific patterns of hydroxymethylation following oxidation and bisulfite treatment (Stewart et al. 2015; Qui et al. 2015).

Single-molecule real-time sequencing on the Pacific Biosciences sequencer shows slight variations in the kinetics of the incorporation of nucleotides depending on DNA base modifications, and the SMRT technology has been used to directly differentiate 5hmc from 5mC and unmodified cytosines (see also Sect. 15.11) (Flusberg et al. 2010). As genome-wide analyses of methylation and hydroxymethylation by SMRT sequencing are not yet feasible (the current output of a SMRT cell is ~ 500 MB), SMRT sequencing has been combined with the above-described glycosylation-mediated enrichment to determine the localization of 5hmC in about 150 MB of sequence (Song et al. 2012).

Hydroxymethylation can be further oxidized to 5-formylcytosine and subsequently to 5-carboxylcytosine by the TET enzymes. Recently, reduced bisulfite sequencing (redBS-seq) was devised, enabling the genome-wide identification of

5-formylcytosine at single-base resolution using the chemical reduction of 5-formylcytosine to 5-hydroxymethylcytosine (Booth et al. 2014). Similar to the above-described OxBS approach, the level of 5fC is obtained by subtraction of the methylation level at cytosines detected in standard BS-seq, which includes of course both 5mC and 5hmC marks from the methylation levels obtained by redBS-seq. Combination of OxBS-seq, standard WGBS, and redBS-seq allows thus the identification of 5mC, 5hmC, and 5fC in the same sample but requires a significant amount of sequencing. The same chemistry has also been combined with an enzymatic approach using the above-described PvuRts1I restriction enzyme, thus avoiding bisulfite conversion (Sun et al. 2015) and selective chemical labeling, whereby 5hmC is modified prior to the reduction of 5fC and the newly created 5hmC is then glycosylated, biotinylated, and sequenced as described above (Song et al. 2013).

11 Direct Readout of DNA Methylation

The direct readout of CpG methylation has been demonstrated as a proof of principle for two amplification-free single-molecule sequencing technologies. These technologies bear the promise of sequencing longer DNA molecules at a single-molecule level, at lower cost and higher speed than existing methods. They can provide information on DNA methylation, hydroxymethylation, and other DNA modifications in the same experiment, at the same time abolishing some of the biases that are inherent to the second-generation sequencing approaches, such as the GC content bias. The single-molecule real-time sequencer from Pacific Biosciences performs sequencing with an immobilized polymerase at the bottom of zero-mode waveguide wells in zeptoliter volumes, monitoring the incorporation of phospho-linked nucleotides through the detection of fluorescent pulses (Flusberg et al. 2010; Song et al. 2012). Single-molecule real-time sequencing on the Pacific Biosciences sequencer does show characteristic variations in the kinetics of the incorporation of nucleotides depending on the DNA base modifications, and the SMRT technology has been used to directly differentiate 5hmc from 5mC and unmodified cytosines. As the kinetics, i.e., the interval between the end of a sequencing pulse and the beginning of the subsequent sequencing pulse, does also depend on the sequence context, an unmethylated template generated through whole-genome amplification is required to define a baseline. While this first report has raised a great interest, very few other reports of direct detection of methylation or hydroxymethylation in mammalian genomes have been made since the original publication 5 years ago. Due to the particularly low signal-to-noise ratio for 5-methylcytosine compared to other DNA modifications like 6-methyladenine, the deconvolution of the kinetic profile has proven to be much more challenging than initially anticipated, especially for regions with high CG content, which are of course of particular interest for DNA methylation. Nonetheless, SMRT sequencing has been used for the detection of epigenetic modifications in bacteria and prokaryotes, where 5-methylcytosine occurs along with 6-methyladenosine and 4-methylcytosine (Blow et al. 2016). Due to their small genome size, exhaustive coverage of >100 × increases the confidence

in the kinetic data and allows the accurate assessment of the methylation status. In a recent example, SMRT sequencing-based DNA methylation analysis was combined with restriction enzyme-based enrichment of a CGG-repeat region of the human *FMR1* gene causing fragile X syndrome (Pham et al. 2016). The accurate methylation analysis of these repeats which are expanded in disease was hitherto not possible due to lack of technologies providing sufficiently long reads. However, the protocol requires large amounts of starting material, and it is based on a complicated multi-step enrichment procedure for the selection of locus of interest due to the necessity to avoid amplification prior to sequencing. Furthermore, only relative and semiquantitative DNA methylation levels could be obtained, and only one of the sequenced strands gave interpretable results, demonstrating that the method has not yet reached maturity for the analysis of human and other mammalian genomes.

Nanopores are an alternative approach and make use of ionic current spectroscopy. The current passing through such pores while DNA is bound to them is very sensitive to subtle structural changes in the interrogated DNA – such as DNA methylation. Nanopore sequencing has the potential to change profoundly the way DNA methylation is analyzed, as besides the absence of potentially artifact-prone steps, such as bisulfite conversion and PCR amplification, it should allow for the direct detection of DNA methylation and its oxidative derivatives and provide longer sequencing reads at lower cost and higher speed. Both solid-state and mutated MspA nanopores have been used for the detection of methylcytosine and its differentiation from hydroxymethylcytosine (Wanunu et al. 2011; Shim et al. 2013; Laszlo et al. 2013; Manrao et al. 2011). Due to their different polarity, DNA flexibility, and duplex stability, solid-state nanopores are capable of differentiating between cytosine, methylated cytosines, and its oxidative derivatives in synthetic templates (Wanunu et al. 2011). However, depending on the nanopore employed, repeated reads might be required to accurately differentiate between 5mC and 5hmC (Laszlo et al. 2013). Depending on the sequence context surrounding the CpG dinucleotides of interest, error rates between 2 and 12 % were observed, suggesting that accurate methylation estimates could be achieved with read numbers of less than 20 molecules of the same locus (Schreiber et al. 2013). In addition, exonuclease-assisted nanopore sequencing of single molecules, for which a processive exonuclease produces nucleoside monophosphates subsequently presented to an alpha-hemolysin protein nanopore with a cyclodextrin adapter, has been shown to detect cytosine modifications in synthetic templates (Clarke et al. 2009; Wallace et al. 2010). However, these technologies have not yielded sufficiently accurate results beyond synthetic templates.

An alternative strategy to increase the ionic blockage induced by the methylation group is the selective labeling of methylcytosine with MBD1 or Kaiso zinc finger proteins, which form a small non-covalent complex with methylated cytosines (Shim et al. 2013, 2015). The binding of the proteins leads to a threefold increase in blockage of the current compared to unmethylated DNA, allowing both the detection of methylated or unmethylated DNA and a rough quantification of the methylation degree of the analyzed DNA fragment by counting the molecules of bound proteins. Similarly, mercury ions can form a reversible bridge between two mismatched DNA bases

containing a thymine-thymine or thymine-uracil mismatch between the analytical target and a synthetic probe (Kang et al. 2013). When passing through a solid-state (graphene) nanopore, changes in the ion current due to the passage of the mercury ion can be recorded. (Methyl)cytosines are however not able to form this reversible interstrand MercuLock, thereby allowing to distinguish methylated cytosines from cytosines after bisulfite conversion (Kang et al. 2013). As the method requires specific probes to be designed for each target, it is more suited for locus-specific analyses rather than potentially genome-wide analysis, and the use of bisulfite conversion does prohibit the differentiation of methylcytosine and hydroxymethylcytosine.

In summary, all so far reported studies on the analysis of DNA methylation with nanopores have used a single (synthetic) template for their proof-of-principle experiments, being still far from the complexity of a human genome. While first results and proof-of-principle experiments are promising, all these technologies require further optimization before being amenable to a routine use in a research laboratory setting.

12 Combined Analysis of DNA Methylation and Other Epigenetic Modifications

12.1 Histone Modifications

The different layers of epigenetic modifications, posttranscriptional histone modifications, histone variants, and DNA methylation, are closely intertwined and stabilize each other to ensure the faithful propagation of an epigenetic state over time and especially through cell division. Studies analyzing several layers of epigenetic modifications are still scarce and due to the requirement of a large amount of biological material often restricted to cell lines and cellular models. Furthermore, the combination of the data is performed rather at the analysis level by overlaying the profiles obtained in distinct experiments and developing probabilistic models on the occurrence of the epigenetic marks rather than by a direct molecular readout. While many technologies have been developed for the comprehensive analysis of a single type of epigenetic modification, few can address the co-occurrence and interaction of different modifications.

Nucleosomes consist of ~147-bp-long DNA stretches wrapped around an octamer of histone proteins that are connected through "linker DNA." The N-terminal tails of the histone proteins are the targets of many posttranslational modifications. The combinations of the different modifications and their multivalency determine the regulatory landscape of a genomic region and its effect on gene expression, giving rise to the concept of the "histone code" (Ernst and Kellis 2012; Hoffman et al. 2013; Rothbart and Strahl 2014). Chromatin immunoprecipitation (ChIP) (Gilmour and Lis 1984) is a well-established method in cellular biology to study the specific interaction between a protein of interest and genomic DNA and has been extensively used to identify transcription factor binding sites (Gerstein et al. 2012).

Briefly, proteins are cross-linked to the DNA by chemicals (normally formaldehyde) in order to conserve the in vivo chromatin architecture. Chromatin is extracted

and randomly fragmented by sonication into 200–600 base pair fragments. Then, DNA-protein complexes are immunoprecipitated using a specific antibody and protein A/G agarose resin. Finally, covalent cross-links are reversed by heating, and DNA is purified after RNase A and proteinase K treatment. At this point, a small amount of purified DNA is available that can be subsequently analyzed by qPCR for the detection and quantification of the analyzed modification at a locus of interest using specific primers, microarrays (ChIP-on-chip), or NGS (ChIP-seq,) for a genome-wide picture of the DNA-protein binding events (i.e., identification of all binding sites of a transcription factor, mapping of a histone modification on the entire genome at very high resolution) (Barski et al. 2007; Mikkelsen et al. 2007; Gerstein et al. 2012). ChIP-seq has since been performed at production scale in the ENCODe, modENCODE, and Roadmap epigenomics mapping projects and has become the workhorse for the genome-wide mapping of the occupancy of DNA by transcription factors and posttranslationally modified histones.

The combination of ChIP with the locus-specific sequencing technologies enables to determine if a protein is (on average) bound to methylated DNA or unmethylated DNA at a given genomic locus. This approach permits studying a putative methyl-binding protein or the association of a specific histone modification with methylated or unmethylated DNA by analyzing the methylation levels of immunoprecipitated DNA at single-nucleotide resolution using bisulfite conversion and analysis of the ChIPed DNA by pyrosequencing (Moison et al. 2013, 2014; Kagey et al. 2010; Thomson et al. 2010). For a more qualitative analysis, methylated DNA can also be specifically detected combining ChIP with methylation-specific PCR (Zinn et al. 2007). Other studies have combined ChIP with genome-wide sequencing methods to obtain a more comprehensive overview of the DNA methylation status with specific histone modifications (ChIP-BS-seq or BisChIP-seq) (Statham et al. 2012; Brinkman et al. 2012). For this approach, a relatively large amount of cells is used for immunoprecipitation to obtain sufficient material for the bisulfite conversion reaction, as well as the ChIP-seq reaction, or several ChIP reactions are pooled prior to library generation. After library preparation with methylated adaptors, part of the library is bisulfite converted and PCR amplified, and the DNA methylation patterns of the immunoprecipitated fragments are analyzed using second-generation sequencing. This analysis permits the simultaneous analysis of DNA methylation associated with histones marked with a set of specific posttranslational modifications on the same pools of cells, but not on the same nucleosome. The above-described SCAN approach (Cipriany et al. 2010, 2012) allows isolating single methylated or unmethylated DNA molecules and enables characterization of epigenetic states, analyzing the coincidence or absence of DNA methylation and histone modification using fluorescently labeled antibodies against the targeted histone modifications in combination with MBD1 for the detection of DNA methylation (Murphy et al. 2013). H3K9me3 was detected together with DNA methylation on the very same nucleosome, while DNA methylation and H3K27me3 were mutually exclusive under normal physiological conditions but became deregulated in cancer or upon cell transformation (Murphy et al. 2013). The use of quantum dots instead of fluorescent dye and a potential parallelization of the nanofluidic devices will allow a higher

throughput of single-molecule analysis with increased multiplexing capabilities. Furthermore, with the advances in low-input/single-cell sequencing described in Sect. 15.9, sequencing of the molecules with a given combinatorial pattern of epigenetic modifications for their identification will probably become feasible.

12.2 Nucleosome Positioning

Positioning of nucleosomes and remodeling of chromatin play key roles for the coordination of the correct gene expression program. Positioning of nucleosomes depends on (among others) the underlying DNA sequence, ATP-dependent nucleosome remodelers, DNA-binding proteins, the RNA polymerase II transcription machinery, and their interactions. As a result, the core enhancer, promoter, and terminator regions of genes are typically depleted of nucleosomes, whereas most of the genomic DNA is occupied (Struhl and Segal 2013). Consequently, the analysis of chromatin accessibility and nucleosome positioning is essential for the understanding of transcriptional regulation, and it can be used for the analysis and identification of gene regulatory elements and their changes in disease. In most cases, information on nucleosome positioning is obtained by enzymatic digestion (MNAse-seq), chemical cleavage (CC-seq), or immunoprecipitation of chromatin, followed by next-generation sequencing of the resulting DNA fragments or derived from chromatin accessibility profiles obtained with DNAseI-seq or ATAC-seq. While well suited for their purpose, these methods do not yield any information on DNA methylation patterns. The DNA *m*ethyltransferase *a*ccessibility *p*rotocol for *i*ndividual *t*emplates (MAPit) or *n*ucleosome *o*ccupancy and *me*thylome sequencing (NOMe-seq) makes use of DNA methyltransferase footprinting to determine the nucleosome positioning while at the same time retaining the original DNA methylation patterns, thereby enabling the correlated analysis of these two epigenetic hallmarks and the corresponding chromatin configurations (Kelly et al. 2012; Pardo et al. 2011). These approaches use the GpC methyltransferase (M.CviPI) (Xu et al. 1998) in the presence of the universal methylation donor S-adenosyl-L-methionine that methylates accessible GC dinucleotides, i.e., sequences that are not protected by nucleosomes or other tight binding proteins. Methylation at CpG dinucleotides yields information about the DNA methylation patterns, while methylation at GpC dinucleotides informs on the chromatin accessibility at the same locus (Jessen et al. 2004). Locus-specific analyses can subsequently be performed by cloning and sequencing after GC methyltransferase treatment (You et al. 2011; Taberlay et al. 2011; Pardo et al. 2011). Genome-wide analyses can also be performed by ligating adaptors to the fragmented and methylated DNA followed by bisulfite conversion and sequencing (Kelly et al. 2012).

Conclusions

Sequencing-based approaches have revolutionized the analysis of the epigenome allowing the analysis of multiple gene regulatory levels including DNA methylation, coding and noncoding RNA expression and its nascent production, location of RNA polymerases, transcription factors and other DNA-binding proteins,

histone modifications, chromatin accessibility, as well as the spatial organization of the genome. With the large distribution of second-generation sequencers and availability of many commercial kits for library preparation, the challenge has shifted from the data generation to the data analysis. Integration of molecular data with other publically available genomic and epigenomic large-scale data sets to analyze the functional consequences of alterations at one molecular level as well as appropriate statistical analysis to decipher higher-order regulatory circuits and the spatial and temporal organization of the genome is currently the major challenge. While still expensive, the deployment of the new Illumina sequencing platforms (X5 and X10) with the possibility to perform whole-genome bisulfite sequencing will further decrease the cost and turnaround time. In addition, the expected continued improvement in sequencing technology and reduction in sequencing costs will enable the routine analysis of methylomes in combination with genomic data. Furthermore, the possibility to determine the methylome of cell-free circulating DNA might be a powerful tool for the early detection of cancers and determination of their origin but might also be useful for the prediction of treatment response in many complex diseases of several organs without disease-causing mutations, such as autoimmune and inflammatory diseases.

Direct readout technologies of epigenetic modifications have been devised and are actively investigated in many laboratories. While their time to technical maturity is difficult to estimate, they promise further decrease in cost and differentiation of methylcytosine from its oxidative derivatives.

While the implementation of the first DNA methylation-based biomarkers has been significantly slowed down due to the use of methods with varying sensitivity and readouts with arbitrary cutoffs and results that were difficult to be integrated and combined between studies, the recent concentration on a few powerful methods with quantitative and often single-nucleotide resolution will enable a much fast progress for the use of locus-specific DNA methylation technologies for clinical diagnosis and prognosis.

Acknowledgments Work in the laboratory of Jörg Tost is supported by grants from the ANR (ANR-13-EPIG-0003-05 and ANR-13-CESA-0011-05), Aviesan/INSERM (EPIG2014-01 and EPIG2014-18), INCa (PRT-K14-049), the joint CEA-EDF-IRSN program (CP-PHE-102), a Sirius Research Award (UCB Pharma S.A.), and the institutional budget of the CNG.

References

Aberg KA, McClay JL, Nerella S, Xie LY, Clark SL, Hudson AD, et al. MBD-seq as a cost-effective approach for methylome-wide association studies: demonstration in 1500 case – control samples. Epigenomics. 2012;4(6):605–21. doi:10.2217/epi.12.59.

Aberg KA, Xie L, Chan RF, Zhao M, Pandey AK, Kumar G, et al. Evaluation of methyl-binding domain based enrichment approaches revisited. PLoS One. 2015;10(7):e0132205. doi:10.1371/journal.pone.0132205.

Absher DM, Li X, Waite LL, Gibson A, Roberts K, Edberg J, et al. Genome-wide DNA methylation analysis of systemic lupus erythematosus reveals persistent hypomethylation of interferon genes and compositional changes to CD4+ T-cell populations. PLoS Genet. 2013;9(8):e1003678. doi:10.1371/journal.pgen.1003678.

Adams D, Altucci L, Antonarakis SE, Ballesteros J, Beck S, Bird A, et al. BLUEPRINT to decode the epigenetic signature written in blood. Nat Biotechnol. 2012;30(3):224–6. doi:10.1038/nbt.2153.

Adey A, Shendure J. Ultra-low-input, tagmentation-based whole-genome bisulfite sequencing. Genome Res. 2012;22(6):1139–43. doi:10.1101/gr.136242.111.

Adey A, Morrison HG, Asan, Xun X, Kitzman JO, Turner EH, et al. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. Genome Biol. 2010;11(12):R119. doi:10.1186/gb-2010-11-12-r119.

Adusumalli S, Mohd Omar MF, Soong R, Benoukraf T. Methodological aspects of whole-genome bisulfite sequencing analysis. Brief Bioinform. 2015;16(3):369–79. doi:10.1093/bib/bbu016.

Ahmadian A, Ehn M, Hober S. Pyrosequencing: history, biochemistry and future. Clin Chim Acta. 2006;363(1–2):83–94. doi:10.1016/j.cccn.2005.04.038.

Alders M, Bliek J, vd Lip K, vd Bogaard R, Mannens M. Determination of KCNQ1OT1 and H19 methylation levels in BWS and SRS patients using methylation-sensitive high-resolution melting analysis. Eur J Hum Genet. 2009;17(4):467–73.

Allum F, Shao X, Guenard F, Simon MM, Busche S, Caron M, et al. Characterization of functional methylomes by next-generation capture sequencing identifies novel disease-associated variants. Nat Commun. 2015;6:7211. doi:10.1038/ncomms8211.

Alnaes GI, Ronneberg JA, Kristensen VN, Tost J. Heterogeneous DNA methylation patterns in the GSTP1 promoter lead to discordant results between assay technologies and impede its implementation as epigenetic biomarkers in breast cancer. Genes (Basel). 2015;6(3):878–900. doi:10.3390/genes6030878.

Andersen GB, Hager H, Hansen LL, Tost J. Improved reproducibility in genome-wide DNA methylation analysis for PAXgene-fixed samples compared with restored formalin fixation and paraffin-embedding DNA. Anal Biochem. 2014;468C:50–8. doi:10.1016/j.ab.2014.09.012.

Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30(10):1363–9. doi:10.1093/bioinformatics/btu049.

Balch C, Yan P, Craft T, Young S, Skalnik DG, Huang TH, et al. Antimitogenic and chemosensitizing effects of the methylation inhibitor zebularine in ovarian cancer. Mol Cancer Ther. 2005;4(10):1505–14.

Balic M, Pichler M, Strutz J, Heitzer E, Ausch C, Samonigg H, et al. High quality assessment of DNA methylation in archival tissues from colorectal cancer patients using quantitative high-resolution melting analysis. J Mol Diagn. 2009;11(2):102–8.

Ball MP, Li JB, Gao Y, Lee JH, LeProust EM, Park IH, et al. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nat Biotechnol. 2009;27(4):361–8. doi:10.1038/nbt.1533.

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. Cell. 2007;129(4):823–37.

Bastian PJ, Palapattu GS, Lin X, Yegnasubramanian S, Mangold LA, Trock B, et al. Preoperative serum DNA GSTP1 CpG island hypermethylation and the risk of early prostate-specific antigen recurrence following radical prostatectomy. Clin Cancer Res. 2005;11(11):4037–43.

Baylin SB, Jones PA. A decade of exploring the cancer epigenome – biological and translational implications. Nat Rev Cancer. 2011;11(10):726–34. doi:10.1038/nrc3130.

Begum S, Brait M, Dasgupta S, Ostrow KL, Zahurak M, Carvalho AL, et al. An epigenetic marker panel for detection of lung cancer using cell-free serum DNA. Clin Cancer Res. 2011;17(13):4494–503. doi:10.1158/1078-0432.CCR-10-3436.

Berdasco M, Fraga MF, Esteller M. Quantification of global DNA methylation by capillary electrophoresis and mass spectrometry. Methods Mol Biol. 2009;507:23–34. doi:10.1007/978-1-59745-522-0_2.

Bestor TH, Ingram VM. Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. Proc Natl Acad Sci U S A. 1983;80(18):5559–63.

Bhattacharyya S, Yu Y, Suzuki M, Campbell N, Mazdo J, Vasanthakumar A, et al. Genome-wide hydroxymethylation tested using the HELP-GT assay shows redistribution in cancer. Nucleic Acids Res. 2013;41(16):e157. doi:10.1093/nar/gkt601.

Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, et al. Genome-wide DNA methylation profiling using Infinium((R)) assay. Epigenomics. 2009;1(1):177–200. doi:10.2217/epi.09.14.

Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. Genomics. 2011;98(4):288–95. doi:10.1016/j.ygeno.2011.07.007. S0888-7543(11)00180-7 [pii].

Bird A. DNA methylation patterns and epigenetic memory. Genes Dev. 2002;16(1):6–21.

Bird AP, Southern EM. Use of restriction enzymes to study eukaryotic DNA methylation: I. The methylation pattern in ribosomal DNA from Xenopus laevis. J Mol Biol. 1978;118(1):27–47.

Blow MJ, Clark TA, Daum CG, Deutschbauer AM, Fomenkov A, Fries R, et al. The epigenomic landscape of prokaryotes. PLoS Genet. 2016;12(2):e1005854. doi:10.1371/journal.pgen.1005854.

Bock C, Tomazou EM, Brinkman AB, Muller F, Simmer F, Gu H, et al. Quantitative comparison of genome-wide DNA methylation mapping technologies. Nat Biotechnol. 2010;28(10):1106–14. doi:10.1038/nbt.1681. nbt.1681 [pii].

Bollati V, Baccarelli A, Hou L, Bonzini M, Fustinoni S, Cavallo D, et al. Changes in DNA methylation patterns in subjects exposed to low-dose benzene. Cancer Res. 2007;67(3):876–80.

Bonanno C, Shehi E, Adlerstein D, Makrigiorgos GM. MS-FLAG, a novel real-time signal generation method for methylation-specific PCR. Clin Chem. 2007;53(12):2119–27. doi:10.1373/clinchem.2007.094011.

Bönsch D, Lenz B, Reulbach U, Kornhuber J, Bleich S. Homocysteine associated genomic DNA hypermethylation in patients with chronic alcoholism. J Neural Transm. 2004;111(12):1611–6.

Booth MJ, Branco MR, Ficz G, Oxley D, Krueger F, Reik W, et al. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. Science. 2012;336(6083):934–7. doi:10.1126/science.1220671.

Booth MJ, Ost TW, Beraldi D, Bell NM, Branco MR, Reik W, et al. Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. Nat Protoc. 2013;8(10):1841–51. doi:10.1038/nprot.2013.115.

Booth MJ, Marsico G, Bachman M, Beraldi D, Balasubramanian S. Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution. Nat Chem. 2014;6(5):435–40. doi:10.1038/nchem.1893.

Borgaro JG, Zhu Z. Characterization of the 5-hydroxymethylcytosine-specific DNA restriction endonucleases. Nucleic Acids Res. 2013;41(7):4198–206. doi:10.1093/nar/gkt102.

Bormann Chung CA, Boyd VL, McKernan KJ, Fu Y, Monighetti C, Peckham HE, et al. Whole methylome analysis by ultra-deep sequencing using two-base encoding. PLoS One. 2010;5(2):e9320. doi:10.1371/journal.pone.0009320.

Borno ST, Fischer A, Kerick M, Falth M, Laible M, Brase JC, et al. Genome-wide DNA methylation events in TMPRSS2-ERG fusion-negative prostate cancers implicate an EZH2-dependent mechanism with miR-26a hypermethylation. Cancer Discov. 2012;2(11):1024–35. doi:10.1158/2159-8290.CD-12-0041.

Brena RM, Auer H, Kornacker K, Hackanson B, Raval A, Byrd JC, et al. Accurate quantification of DNA methylation using combined bisulfite restriction analysis coupled with the Agilent 2100 Bioanalyzer platform. Nucleic Acids Res. 2006;34(3):e17.

Brinkman AB, Simmer F, Ma K, Kaan A, Zhu J, Stunnenberg HG. Whole-genome DNA methylation profiling using MethylCap-seq. Methods. 2010;52(3):232–6. doi:10.1016/j.ymeth.2010.06.012. S1046-2023(10)00166-0 [pii].

Brinkman AB, Gu H, Bartels SJ, Zhang Y, Matarese F, Simmer F, et al. Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. Genome Res. 2012;22(6):1128–38. doi:10.1101/gr.133728.111.

Brouilette S, Kuersten S, Mein C, Bozek M, Terry A, Dias KR, et al. A simple and novel method for RNA-seq library preparation of single cell cDNA analysis by hyperactive Tn5 transposase. Dev Dyn. 2012;241(10):1584–90. doi:10.1002/dvdy.23850.

Brunner AL, Johnson DS, Kim SW, Valouev A, Reddy TE, Neff NF, et al. Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver. Genome Res. 2009;19(6):1044–56. doi:10.1101/gr.088773.108.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013;10(12):1213–8. doi:10.1038/nmeth.2688.

Bundo M, Sunaga F, Ueda J, Kasai K, Kato T, Iwamoto K. A systematic evaluation of whole genome amplification of bisulfite-modified DNA. Clini Epigenetics. 2012;4(1):22. doi:10.1186/1868-7083-4-22.

Campan M, Weisenberger DJ, Trinh B, Laird PW. MethyLight. Methods Mol Biol. 2009;507:325–37. doi:10.1007/978-1-59745-522-0_23.

Campan M, Moffitt M, Houshdaran S, Shen H, Widschwendter M, Daxenbichler G, et al. Genome-scale screen for DNA methylation-based detection markers for ovarian cancer. PLoS One. 2011;6(12):e28141. doi:10.1371/journal.pone.0028141. PONE-D-11-08732 [pii].

Candiloro IL, Mikeska T, Dobrovic A. Assessing combined methylation-sensitive high resolution melting and pyrosequencing for the analysis of heterogeneous DNA methylation. Epigenetics. 2011;6(4):500–7.

Castellanos-Rizaldos E, Milbury CA, Karatza E, Chen CC, Makrigiorgos GM, Merewood A. COLD-PCR amplification of bisulfite-converted DNA allows the enrichment and sequencing of rare un-methylated genomic regions. PLoS One. 2014;9(4):e94103. doi:10.1371/journal.pone.0094103.

Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, et al. Plant genetics. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. Science. 2014;345(6199):950–3. doi:10.1126/science.1253435.

Chan KC, Jiang P, Chan CW, Sun K, Wong J, Hui EP, et al. Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing. Proc Natl Acad Sci U S A. 2013a;110(47):18761–8. doi:10.1073/pnas.1313995110.

Chan KC, Jiang P, Zheng YW, Liao GJ, Sun H, Wong J, et al. Cancer genome scanning in plasma: detection of tumor-associated copy number aberrations, single-nucleotide variants, and tumoral heterogeneity by massively parallel sequencing. Clin Chem. 2013b;59(1):211–24. doi:10.1373/clinchem.2012.196014.

Chan RW, Jiang P, Peng X, Tam LS, Liao GJ, Li EK, et al. Plasma DNA aberrations in systemic lupus erythematosus revealed by genomic and methylomic sequencing. Proc Natl Acad Sci U S A. 2014;111(49):E5302–11. doi:10.1073/pnas.1421126111.

Chandra A, Ray A, Senapati S, Chatterjee R. Genetic and epigenetic basis of psoriasis pathogenesis. Mol Immunol. 2015;64(2):313–23. doi:10.1016/j.molimm.2014.12.014.

Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics. 2013;8(2):203–9. doi:10.4161/epi.23470.

Cheow LF, Quake SR, Burkholder WF, Messerschmidt DM. Multiplexed locus-specific analysis of DNA methylation in single cells. Nat Protoc. 2015;10(4):619–31. doi:10.1038/nprot.2015.041.

Chhibber A, Schroeder BG. Single-molecule polymerase chain reaction reduces bias: application to DNA methylation analysis by bisulfite sequencing. Anal Biochem. 2008;377(1):46–54. doi:10.1016/j.ab.2008.02.026.

Chim SS, Tong YK, Chiu RW, Lau TK, Leung TN, Chan LY, et al. Detection of the placental epigenetic signature of the maspin gene in maternal plasma. Proc Natl Acad Sci U S A. 2005;102(41):14753–8. doi:10.1073/pnas.0503335102. 0503335102 [pii].

Chim SS, Jin S, Lee TY, Lun FM, Lee WS, Chan LY, et al. Systematic search for placental DNA-methylation markers on chromosome 21: toward a maternal plasma-based epigenetic test

for fetal trisomy 21. Clin Chem. 2008;54(3):500–11. doi:10.1373/clinchem.2007.098731. clinchem.2007.098731 [pii].

Choi JH, Li Y, Guo J, Pei L, Rauch TA, Kramer RS, et al. Genome-wide DNA methylation maps in follicular lymphoma cells determined by methylation-enriched bisulfite sequencing. PLoS One. 2010;5(9). doi:10.1371/journal.pone.0013020.

Chopra P, Papale LA, White AT, Hatch A, Brown RM, Garthwaite MA, et al. Array-based assay detects genome-wide 5-mC and 5-hmC in the brains of humans, non-human primates, and mice. BMC Genomics. 2014;15:131. doi:10.1186/1471-2164-15-131.

Chu T, Burke B, Bunce K, Surti U, Allen Hogge W, Peters DG. A microarray-based approach for the identification of epigenetic biomarkers for the noninvasive diagnosis of fetal disease. Prenat Diagn. 2009;29(11):1020–30. doi:10.1002/pd.2335.

Church TR, Wandell M, Lofton-Day C, Mongin SJ, Burger M, Payne SR, et al. Prospective evaluation of methylated SEPT9 in plasma for detection of asymptomatic colorectal cancer. Gut. 2014;63(2):317–25. doi:10.1136/gutjnl-2012-304149.

Cipriany BR, Zhao R, Murphy PJ, Levy SL, Tan CP, Craighead HG, et al. Single molecule epigenetic analysis in a nanofluidic channel. Anal Chem. 2010;82(6):2480–7. doi:10.1021/ac9028642.

Cipriany BR, Murphy PJ, Hagarman JA, Cerf A, Latulippe D, Levy SL, et al. Real-time analysis and selection of methylated DNA by fluorescence-activated single molecule sorting in a nanofluidic channel. Proc Natl Acad Sci U S A. 2012;109(22):8477–82. doi:10.1073/pnas.1117549109.

Clark C, Palta P, Joyce CJ, Scott C, Grundberg E, Deloukas P, et al. A comparison of the whole genome approach of MeDIP-seq to the targeted approach of the Infinium HumanMethylation450 BeadChip((R)) for methylome profiling. PLoS One. 2012;7(11):e50233. doi:10.1371/journal.pone.0050233.

Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H. Continuous base identification for single-molecule nanopore DNA sequencing. Nat Nanotechnol. 2009;4(4):265–70. doi:10.1038/nnano.2009.12. nnano.2009.12 [pii].

Claus R, Wilop S, Hielscher T, Sonnet M, Dahl E, Galm O, et al. A systematic comparison of quantitative high-resolution DNA methylation analysis and methylation-specific PCR. Epigenetics. 2012;7(7):772–80. doi:10.4161/epi.20299.

Cohen-Karni D, Xu D, Apone L, Fomenkov A, Sun Z, Davis PJ, et al. The MspJI family of modification-dependent restriction endonucleases for epigenetic studies. Proc Natl Acad Sci U S A. 2011;108(27):11040–5. doi:10.1073/pnas.1018448108.

Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, et al. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature. 2008;452(7184):215–9.

Colaneri A, Staffa N, Fargo DC, Gao Y, Wang T, Peddada SD, et al. Expanded methyl-sensitive cut counting reveals hypomethylation as an epigenetic state that highlights functional sequences of the genome. Proc Natl Acad Sci U S A. 2011;108(23):9715–20. doi:10.1073/pnas.1105713108.

Coolen MW, Statham AL, Gardiner-Garden M, Clark SJ. Genomic profiling of CpG methylation and allelic specificity using quantitative high-throughput mass spectrometry: critical evaluation and improvements. Nucleic Acids Res. 2007;35(18):e119. doi:10.1093/nar/gkm662.

Corley MJ, Zhang W, Zheng X, Lum-Jones A, Maunakea AK. Semiconductor-based sequencing of genome-wide DNA methylation states. Epigenetics. 2015;10(2):153–66. doi:10.1080/15592294.2014.1003747.

Cottrell SE, Distler J, Goodman NS, Mooney SH, Kluth A, Olek A, et al. A real-time PCR assay for DNA-methylation using methylation-specific blockers. Nucleic Acids Res. 2004;32(1):e10.

Crowley E, Di Nicolantonio F, Loupakis F, Bardelli A. Liquid biopsy: monitoring cancer-genetics in the blood. Nat Rev Clin Oncol. 2013;10(8):472–84. doi:10.1038/nrclinonc.2013.110.

Cui L, Chung TH, Tan D, Sun X, Jia XY. JBP1-seq: a fast and efficient method for genome-wide profiling of 5hmC. Genomics. 2014;104(5):368–75. doi:10.1016/j.ygeno.2014.08.023.

Daugaard I, Kjeldsen TE, Hager H, Hansen LL, Wojdacz TK. The influence of DNA degradation in formalin-fixed, paraffin-embedded (FFPE) tissue on locus-specific methylation assessment by MS-HRM. Exp Mol Pathol. 2015;99(3):632–40. doi:10.1016/j.yexmp.2015.11.007.

De Meyer T, Mampaey E, Vlemmix M, Denil S, Trooskens G, Renard JP, et al. Quality evaluation of methyl binding domain based kits for enrichment DNA-methylation sequencing. PLoS One. 2013;8(3):e59068. doi:10.1371/journal.pone.0059068.

Delatte B, Deplus R, Fuks F. Playing TETris with DNA modifications. EMBO J. 2014;33(11):1198–211. doi:10.15252/embj.201488290.

Deng J, Shoemaker R, Xie B, Gore A, LeProust EM, Antosiewicz-Bourget J, et al. Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. Nat Biotechnol. 2009;27(4):353–60. doi:10.1038/nbt.1530.

Diaz Jr LA, Bardelli A. Liquid biopsies: genotyping circulating tumor DNA. J Clin Oncol. 2014;32(6):579–86. doi:10.1200/JCO.2012.45.2011.

Dick KJ, Nelson CP, Tsaprouni L, Sandling JK, Aissi D, Wahl S, et al. DNA methylation and body-mass index: a genome-wide analysis. Lancet. 2014;383(9933):1990–8. doi:10.1016/S0140-6736(13)62674-4.

Diep D, Plongthongkum N, Gore A, Fung HL, Shoemaker R, Zhang K. Library-free methylation sequencing with bisulfite padlock probes. Nat Methods. 2012;9(3):270–2. doi:10.1038/nmeth.1871.

Dikow N, Nygren AO, Schouten JP, Hartmann C, Kramer N, Janssen B, et al. Quantification of the methylation status of the PWS/AS imprinted region: comparison of two approaches based on bisulfite sequencing and methylation-sensitive MLPA. Mol Cell Probes. 2007;21(3):208–15.

Down TA, Rakyan VK, Turner DJ, Flicek P, Li H, Kulesha E, et al. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. Nat Biotechnol. 2008;26(7):779–85.

Dumenil TD, Wockner LF, Bettington M, McKeone DM, Klein K, Bowdler LM, et al. Genome-wide DNA methylation analysis of formalin-fixed paraffin embedded colorectal cancer tissue. Genes Chromosomes Cancer. 2014;53(7):537–48. doi:10.1002/gcc.22164.

Dupont JM, Tost J, Jammes H, Gut IG. De novo quantitative bisulfite sequencing using the pyrosequencing technology. Anal Biochem. 2004;333(1):119–27. doi:10.1016/j.ab.2004.05.007. S0003269704004130 [pii].

Eads CA, Danenberg KD, Kawakami K, Saltz LB, Blake C, Shibata D, et al. MethyLight: a high-throughput assay to measure DNA methylation. Nucleic Acids Res. 2000;28(8):e32.

Edwards JR, O'Donnell AH, Rollins RA, Peckham HE, Lee C, Milekic MH, et al. Chromatin and sequence features that define the fine and gross structure of genomic methylation patterns. Genome Res. 2010;20(7):972–80. doi:10.1101/gr.101535.109.

Ehrich M, Nelson MR, Stanssens P, Zabeau M, Liloglou T, Xinarianos G, et al. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. Proc Natl Acad Sci U S A. 2005;102(44):15785–90. doi:10.1073/pnas.0507816102.

Ehrich M, Zoll S, Sur S, van den Boom D. A new method for accurate assessment of DNA quality after bisulfite treatment. Nucleic Acids Res. 2007;35(5):e29. doi:10.1093/nar/gkl1134. gkl1134 [pii].

Eick D, Fritz HJ, Doerfler W. Quantitative determination of 5-methylcytosine in DNA by reverse-phase high-performance liquid chromatography. Anal Biochem. 1983;135(1):165–71.

Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nat Methods. 2012;9(3):215–6. doi:10.1038/nmeth.1906.

Farlik M, Sheffield NC, Nuzzo A, Datlinger P, Schonegger A, Klughammer J, et al. Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. Cell Rep. 2015;10(8):1386–97. doi:10.1016/j.celrep.2015.02.001.

Fazzari MJ, Greally JM. Epigenomics: beyond CpG islands. Nat Rev Genet. 2004;5(6):446–55.

Feber A, Wilson GA, Zhang L, Presneau N, Idowu B, Down TA, et al. Comparative methylome analysis of benign and malignant peripheral nerve sheath tumors. Genome Res. 2011;21(4):515–24. doi:10.1101/gr.109678.110.

Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, et al. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature. 2011;473(7347):398–402. doi:10.1038/nature10008.

Field SF, Beraldi D, Bachman M, Stewart SK, Beck S, Balasubramanian S. Accurate measurement of 5-methylcytosine and 5-hydroxymethylcytosine in human cerebellum DNA by oxidative bisulfite on an array (OxBS-array). PLoS One. 2015;10(2):e0118202. doi:10.1371/journal.pone.0118202.

Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods. 2010;7(6):461–5. doi:10.1038/nmeth.1459. nmeth.1459 [pii].

Fraga MF, Uriol E, Borja Diego L, Berdasco M, Esteller M, Canal MJ, et al. High-performance capillary electrophoretic method for the quantification of 5-methyl 2′-deoxycytidine in genomic DNA: application to plant, animal and human cancer tissues. Electrophoresis. 2002;23(11):1677–81.

Fraga MF, Ballestar E, Montoya G, Taysavang P, Wade PA, Esteller M. The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. Nucleic Acids Res. 2003;31(6):1765–74.

Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci U S A. 1992;89(5):1827–31.

Gebhard C, Schwarzfischer L, Pham TH, Schilling E, Klug M, Andreesen R, et al. Genome-wide profiling of CpG methylation identifies novel targets of aberrant hypermethylation in myeloid leukemia. Cancer Res. 2006;66(12):6118–28. doi:10.1158/0008-5472.CAN-06-0376.

Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, Cheng C, et al. Architecture of the human regulatory network derived from ENCODE data. Nature. 2012;489(7414):91–100. doi:10.1038/nature11245.

Ghantous A, Saffery R, Cros MP, Ponsonby AL, Hirschfeld S, Kasten C, et al. Optimized DNA extraction from neonatal dried blood spots: application in methylome profiling. BMC Biotechnol. 2014;14:60. doi:10.1186/1472-6750-14-60.

Gilmour DS, Lis JT. Detecting protein-DNA interactions in vivo: distribution of RNA polymerase on specific bacterial genes. Proc Natl Acad Sci U S A. 1984;81(14):4275–9.

Globisch D, Munzel M, Muller M, Michalakis S, Wagner M, Koch S, et al. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PLoS One. 2010;5(12):e15367. doi:10.1371/journal.pone.0015367.

Glossop JR, Emes RD, Nixon NB, Haworth KE, Packham JC, Dawes PT, et al. Genome-wide DNA methylation profiling in rheumatoid arthritis identifies disease-associated methylation changes that are distinct to individual T- and B-lymphocyte populations. Epigenetics. 2014;9(9):1228–37. doi:10.4161/epi.29718.

Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol. 2009;27(2):182–9. doi:10.1038/nbt.1523.

Godderis L, Schouteden C, Tabish A, Poels K, Hoet P, Baccarelli AA, et al. Global methylation and hydroxymethylation in DNA from blood and Saliva in Healthy Volunteers. Biomed Res Int. 2015;2015:845041. doi:10.1155/2015/845041.

Gonzalgo ML, Jones PA. Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE). Nucleic Acids Res. 1997;25(12):2529–31.

Gravina S, Ganapathi S, Vijg J. Single-cell, locus-specific bisulfite sequencing (SLBS) for direct detection of epimutations in DNA methylation patterns. Nucleic Acids Res. 2015;43(14):e93. doi:10.1093/nar/gkv366.

Gries J, Schumacher D, Arand J, Lutsik P, Markelova MR, Fichtner I, et al. Bi-PROF: bisulfite profiling of target regions using 454 GS FLX Titanium technology. Epigenetics. 2013;8(7):765–71. doi:10.4161/epi.25242.

Gross JA, Pacis A, Chen GG, Barreiro LB, Ernst C, Turecki G. Characterizing 5-hydroxymethylcytosine in human prefrontal cortex at single base resolution. BMC Genomics. 2015;16:672. doi:10.1186/s12864-015-1875-8.

Grunau C, Clark SJ, Rosenthal A. Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. Nucleic Acids Res. 2001;29(13):e65.

Grundberg E, Meduri E, Sandling JK, Hedman AK, Keildson S, Buil A, et al. Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. Am J Hum Genet. 2013;93(5):876–90. doi:10.1016/j.ajhg.2013.10.004.

Gu H, Smith ZD, Bock C, Boyle P, Gnirke A, Meissner A. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. Nat Protoc. 2011;6(4):468–81. doi:10.1038/nprot.2010.190.

Guo H, Zhu P, Wu X, Li X, Wen L, Tang F. Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. Genome Res. 2013;23(12):2126–35. doi:10.1101/gr.161679.113.

Guo JU, Su Y, Shin JH, Shin J, Li H, Xie B, et al. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. Nat Neurosci. 2014;17(2):215–22. doi:10.1038/nn.3607.

Guo H, Zhu P, Guo F, Li X, Wu X, Fan X, et al. Profiling DNA methylome landscapes of mammalian cells with single-cell reduced-representation bisulfite sequencing. Nat Protoc. 2015;10(5):645–59. doi:10.1038/nprot.2015.039.

Gupta S, Jaworska-Bieniek K, Narod SA, Lubinski J, Wojdacz TK, Jakubowska A. Methylation of the BRCA1 promoter in peripheral blood DNA is associated with triple-negative and medullary breast cancer. Breast Cancer Res Treat. 2014;148(3):615–22. doi:10.1007/s10549-014-3179-0.

Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, et al. Increased methylation variation in epigenetic domains across cancer types. Nat Genet. 2011;43(8):768–75. doi:10.1038/ng.865.

Hardy T, Zeybel M, Day CP, Dipper C, Masson S, McPherson S, et al. Plasma DNA methylation: a potential biomarker for stratification of liver fibrosis in non-alcoholic fatty liver disease. Gut. 2016. doi:10.1136/gutjnl-2016-311526.

Harrington CT, Lin EI, Olson MT, Eshleman JR. Fundamentals of pyrosequencing. Arch Pathol Lab Med. 2013;137(9):1296–303. doi:10.5858/arpa.2012-0463-RA.

Harris RA, Wang T, Coarfa C, Nagarajan RP, Hong C, Downey SL, et al. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. Nat Biotechnol. 2010;28(10):1097–105. doi:10.1038/nbt.1682.

Harrison A, Parle-McDermott A. DNA methylation: a timeline of methods and applications. Front Genet. 2011;2:74. doi:10.3389/fgene.2011.00074.

Heitzer E, Ulz P, Belic J, Gutschi S, Quehenberger F, Fischereder K, et al. Tumor-associated copy number changes in the circulation of patients with prostate cancer identified through whole-genome sequencing. Genome Med. 2013;5(4):30. doi:10.1186/gm434.

Heitzer E, Ulz P, Geigl JB. Circulating tumor DNA as a liquid biopsy for cancer. Clin Chem. 2015;61(1):112–23. doi:10.1373/clinchem.2014.222679.

Henkhaus RS, Kim SJ, Kimonis VE, Gold JA, Dykens EM, Driscoll DJ, et al. Methylation-specific multiplex ligation-dependent probe amplification and identification of deletion genetic subtypes in Prader-Willi syndrome. Genet Test Mol Biomarkers. 2012;16(3):178–86. doi:10.1089/gtmb.2011.0115.

Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. Proc Natl Acad Sci U S A. 1996;93(18):9821–6.

Hodges E, Smith AD, Kendall J, Xuan Z, Ravi K, Rooks M, et al. High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. Genome Res. 2009;19(9):1593–605. doi:10.1101/gr.095190.109.

Hoffman MM, Ernst J, Wilder SP, Kundaje A, Harris RS, Libbrecht M, et al. Integrative annotation of chromatin elements from ENCODE data. Nucleic Acids Res. 2013;41(2):827–41. doi:10.1093/nar/gks1284.

Holland PM, Abramson RD, Watson R, Gelfand DH. Detection of specific polymerase chain reaction product by utilizing the 5′–3′ exonuclease activity of Thermus aquaticus DNA polymerase. Proc Natl Acad Sci U S A. 1991;88(16):7276–80.

Holmes EE, Jung M, Meller S, Leisse A, Sailer V, Zech J, et al. Performance evaluation of kits for bisulfite-conversion of DNA from tissues, cell lines, FFPE tissues, aspirates, lavages, effusions, plasma, serum, and urine. PLoS One. 2014;9(4):e93933. doi:10.1371/journal.pone.0093933.
Homig-Holzel C, Savola S. Multiplex ligation-dependent probe amplification (MLPA) in tumor diagnostics and prognostics. Diagn Mol Pathol. 2012;21(4):189–206. doi:10.1097/PDM.0b013e3182595516.
Hoon DS, Spugnardi M, Kuo C, Huang SK, Morton DL, Taback B. Profiling epigenetic inactivation of tumor suppressor genes in tumors and plasma from cutaneous melanoma patients. Oncogene. 2004;23(22):4014–22. doi:10.1038/sj.onc.1207505.
Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012;13:86. doi:10.1186/1471-2105-13-86.
How Kit A, Nielsen HM, Tost J. DNA methylation based biomarkers: practical considerations and applications. Biochimie. 2012;94(11):2314–37. doi:10.1016/j.biochi.2012.07.014.
Hu CY, Mohtat D, Yu Y, Ko YA, Shenoy N, Bhattacharya S, et al. Kidney cancer is characterized by aberrant methylation of tissue-specific enhancers that are prognostic for overall survival. Clin Cancer Res. 2014;20(16):4349–60. doi:10.1158/1078-0432.CCR-14-0494.
Huang TH, Perry MR, Laux DE. Methylation profiling of CpG islands in human breast cancer cells. Hum Mol Genet. 1999;8(3):459–70.
Huang DJ, Nelson MR, Zimmermann B, Dudarewicz L, Wenzel F, Spiegel R, et al. Reliable detection of trisomy 21 using MALDI-TOF mass spectrometry. Genet Med. 2006;8(11):728–34. doi:10.1097/01.gim.0000245573.42908.34. 00125817-200611000-00008 [pii].
Huang J, Renault V, Sengenes J, Touleimat N, Michel S, Lathrop M, et al. MeQA: a pipeline for MeDIP-seq data quality assessment and analysis. Bioinformatics. 2012;28(4):587–8. doi:10.1093/bioinformatics/btr699.
Hur K, Cejas P, Feliu J, Moreno-Rubio J, Burgos E, Boland CR, et al. Hypomethylation of long interspersed nuclear element-1 (LINE-1) leads to activation of proto-oncogenes in human colorectal cancer metastasis. Gut. 2014;63(4):635–46. doi:10.1136/gutjnl-2012-304219.
Ilse P, Biesterfeld S, Pomjanski N, Wrobel C, Schramm M. Analysis of SHOX2 methylation as an aid to cytology in lung cancer diagnosis. Cancer Genomics Proteomics. 2014;11(5):251–8.
Inamura K, Yamauchi M, Nishihara R, Lochhead P, Qian ZR, Kuchiba A, et al. Tumor LINE-1 methylation level and microsatellite instability in relation to colorectal cancer prognosis. J Natl Cancer Inst. 2014;106(9). doi:10.1093/jnci/dju195.
Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet. 2009;41(2):178–86. doi:10.1038/ng.298. ng.298 [pii].
Ivanov M, Kals M, Kacevska M, Metspalu A, Ingelman-Sundberg M, Milani L. In-solution hybrid capture of bisulfite-converted DNA for targeted bisulfite sequencing of 174 ADME genes. Nucleic Acids Res. 2013;41(6):e72. doi:10.1093/nar/gks1467.
Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, et al. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science. 2014;343(6172):776–9. doi:10.1126/science.1247651.
Jasmine F, Rahaman R, Roy S, Raza M, Paul R, Rakibuz-Zaman M, et al. Interpretation of genome-wide infinium methylation data from ligated DNA in formalin-fixed, paraffin-embedded paired tumor and normal tissue. BMC Res Notes. 2012;5:117. doi:10.1186/1756-0500-5-117.
Jelinek J, Liang S, Lu Y, He R, Ramagli LS, Shpall EJ, et al. Conserved DNA methylation patterns in healthy blood cells and extensive changes in leukemia measured by a new quantitative technique. Epigenetics. 2012;7(12):1368–78. doi:10.4161/epi.22552.
Jenkins TG, Aston KI, Pflueger C, Cairns BR, Carrell DT. Age-associated sperm DNA methylation alterations: possible implications in offspring disease susceptibility. PLoS Genet. 2014;10(7):e1004458. doi:10.1371/journal.pgen.1004458.
Jessen WJ, Dhasarathy A, Hoose SA, Carvin CD, Risinger AL, Kladde MP. Mapping chromatin structure in vivo using DNA methyltransferases. Methods. 2004;33(1):68–80. doi:10.1016/j.ymeth.2003.10.025.

Jin SG, Kadam S, Pfeifer GP. Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. Nucleic Acids Res. 2010;38(11):e125. doi:10.1093/nar/gkq223.

Jung M, Kadam S, Xiong W, Rauch TA, Jin SG, Pfeifer GP. MIRA-seq for DNA methylation analysis of CpG islands. Epigenomics. 2015;7(5):695–706. doi:10.2217/epi.15.33.

Kagey JD, Kapoor-Vazirani P, McCabe MT, Powell DR, Vertino PM. Long-term stability of demethylation after transient exposure to 5-aza-2′-deoxycytidine correlates with sustained RNA polymerase II occupancy. Mol Cancer Res. 2010;8(7):1048–59. doi:10.1158/1541-7786.MCR-10-0189.

Kaminsky ZA, Assadzadeh A, Flanagan J, Petronis A. Single nucleotide extension technology for quantitative site-specific evaluation of metC/C in GC-rich regions. Nucleic Acids Res. 2005;33(10):e95.

Kang I, Wang Y, Reagan C, Fu Y, Wang MX, Gu LQ. Designing DNA interstrand lock for locus-specific methylation detection in a nanopore. Sci Rep. 2013;3:2381. doi:10.1038/srep02381.

Kantarjian H, Issa JP, Rosenfeld CS, Bennett JM, Albitar M, DiPersio J, et al. Decitabine improves patient outcomes in myelodysplastic syndromes: results of a phase III randomized study. Cancer. 2006;106(8):1794–803.

Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10000 daltons. Anal Chem. 1988;60:2299–301.

Karatzas PS, Gazouli M, Safioleas M, Mantzaris GJ. DNA methylation changes in inflammatory bowel disease. Ann Gastroenterol. 2014;27(2):125–32.

Karimi M, Johansson S, Stach D, Corcoran M, Grander D, Schalling M, et al. LUMA (LUminometric Methylation Assay) – a high throughput method to the analysis of genomic DNA methylation. Exp Cell Res. 2006;312(11):1989–95.

Kelly TK, Liu Y, Lay FD, Liang G, Berman BP, Jones PA. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. Genome Res. 2012;22(12):2497–506. doi:10.1101/gr.143008.112.

Khulan B, Thompson RF, Ye K, Fazzari MJ, Suzuki M, Stasiek E, et al. Comparative isoschizomer profiling of cytosine methylation: the HELP assay. Genome Res. 2006;16(8):1046–55.

Kinney SM, Chin HG, Vaisvila R, Bitinaite J, Zheng Y, Esteve PO, et al. Tissue-specific distribution and dynamic changes of 5-hydroxymethylcytosine in mammalian genomes. J Biol Chem. 2011;286(28):24685–93. doi:10.1074/jbc.M110.217083.

Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell. 2015;161(5):1187–201. doi:10.1016/j.cell.2015.04.044.

Kobayashi H, Sakurai T, Imai M, Takahashi N, Fukuda A, Yayoi O, et al. Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. PLoS Genet. 2012;8(1):e1002440. doi:10.1371/journal.pgen.1002440.

Kobayashi H, Sakurai T, Miura F, Imai M, Mochiduki K, Yanagisawa E, et al. High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. Genome Res. 2013;23(4):616–27. doi:10.1101/gr.148023.112.

Komori HK, LaMere SA, Torkamani A, Hart GT, Kotsopoulos S, Warner J, et al. Application of microdroplet PCR for large-scale targeted bisulfite sequencing. Genome Res. 2011;21(10):1738–45. doi:10.1101/gr.116863.110.

Korbie D, Lin E, Wall D, Nair SS, Stirzaker C, Clark SJ, et al. Multiplex bisulfite PCR resequencing of clinical FFPE DNA. Clini Epigenetics. 2015;7(1):28. doi:10.1186/s13148-015-0067-3.

Kreck B, Marnellos G, Richter J, Krueger F, Siebert R, Franke A. B-SOLANA: an approach for the analysis of two-base encoding bisulfite sequencing data. Bioinformatics. 2012;28(3):428–9. doi:10.1093/bioinformatics/btr660.

Kreck B, Richter J, Ammerpohl O, Barann M, Esser D, Petersen BS, et al. Base-pair resolution DNA methylome of the EBV-positive Endemic Burkitt lymphoma cell line DAUDI determined by SOLiD bisulfite-sequencing. Leukemia. 2013;27(8):1751–3. doi:10.1038/leu.2013.4.

Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science. 2009;324(5929):929–30. doi:10.1126/science.1169786. 1169786 [pii].

Kristensen LS, Mikeska T, Krypuy M, Dobrovic A. Sensitive Melting Analysis after Real Time-Methylation Specific PCR (SMART-MSP): high-throughput and probe-free quantitative DNA methylation detection. Nucleic Acids Res. 2008;36(7):e42.

Kristensen LS, Wojdacz TK, Thestrup BB, Wiuf C, Hager H, Hansen LL. Quality assessment of DNA derived from up to 30 years old formalin fixed paraffin embedded (FFPE) tissue for PCR-based methylation analysis using SMART-MSP and MS-HRM. BMC Cancer. 2009;9:453. doi:10.1186/1471-2407-9-453. 1471-2407-9-453 [pii].

Kriukiene E, Labrie V, Khare T, Urbanaviciute G, Lapinaite A, Koncevicius K, et al. DNA unmethylome profiling by covalent capture of CpG sites. Nat Commun. 2013;4:2190. doi:10.1038/ncomms3190.

Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011;27(11):1571–2. doi:10.1093/bioinformatics/btr167.

Laird PW. Early detection: the power and the promise of DNA methylation markers. Nat Rev Cancer. 2003;3(4):253–66.

Lan X, Adams C, Landers M, Dudas M, Krissinger D, Marnellos G, et al. High resolution detection and analysis of CpG dinucleotides methylation using MBD-Seq technology. PLoS One. 2011;6(7):e22226. doi:10.1371/journal.pone.0022226.

Langaee T, Ronaghi M. Genetic variation analyses by Pyrosequencing. Mutat Res. 2005;573(1–2): 96–102.

Laszlo AH, Derrington IM, Brinkerhoff H, Langford KW, Nova IC, Samson JM, et al. Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. Proc Natl Acad Sci U S A. 2013;110(47):18904–9. doi:10.1073/pnas.1310240110.

Le T, Kim KP, Fan G, Faull KF. A sensitive mass spectrometry method for simultaneous quantification of DNA methylation and hydroxymethylation levels in biological samples. Anal Biochem. 2011;412(2):203–9. doi:10.1016/j.ab.2011.01.026.

Lee EJ, Pei L, Srivastava G, Joshi T, Kushwaha G, Choi JH, et al. Targeted bisulfite sequencing by solution hybrid selection and massively parallel sequencing. Nucleic Acids Res. 2011;39(19):e127. doi:10.1093/nar/gkr598.

Lehmann-Werman R, Neiman D, Zemmour H, Moss J, Magenheim J, Vaknin-Dembinsky A, et al. Identification of tissue-specific cell death using methylation patterns of circulating DNA. Proc Natl Acad Sci U S A. 2016;113(13):E1826–34. doi:10.1073/pnas.1519286113.

Leong KJ, James J, Wen K, Taniere P, Morton DG, Bach SP, et al. Impact of tissue processing, archiving and enrichment techniques on DNA methylation yield in rectal carcinoma. Exp Mol Pathol. 2013;95(3):343–9. doi:10.1016/j.yexmp.2013.10.007.

Li J, Gao F, Li N, Li S, Yin G, Tian G, et al. An improved method for genome wide DNA methylation profiling correlated to transcription and genomic instability in two breast cancer cell lines. BMC Genomics. 2009;10:223. doi:10.1186/1471-2164-10-223.

Li N, Ye M, Li Y, Yan Z, Butcher LM, Sun J, et al. Whole genome DNA methylation analysis based on high throughput sequencing technology. Methods. 2010a;52(3):203–12. doi:10.1016/j.ymeth.2010.04.009.

Li Y, Zhu J, Tian G, Li N, Li Q, Ye M, et al. The DNA methylome of human peripheral blood mononuclear cells. PLoS Biol. 2010b;8(11):e1000533. doi:10.1371/journal.pbio.1000533.

Li D, Zhang B, Xing X, Wang T. Combining MeDIP-seq and MRE-seq to investigate genome-wide CpG methylation. Methods. 2015a;72:29–40. doi:10.1016/j.ymeth.2014.10.032.

Li H, Hedmer M, Wojdacz T, Hossain MB, Lindh CH, Tinnerberg H, et al. Oxidative stress, telomere shortening, and DNA methylation in relation to low-to-moderate occupational exposure to welding fumes. Environ Mol Mutagen. 2015b;56(8):684–93. doi:10.1002/em.21958.

Li Q, Suzuki M, Wendt J, Patterson N, Eichten SR, Hermanson PJ, et al. Post-conversion targeted capture of modified cytosines in mammalian and plant genomes. Nucleic Acids Res. 2015c;43(12):e81. doi:10.1093/nar/gkv244.

Li H, Akerman G, Liden C, Alhamdow A, Wojdacz TK, Broberg K, et al. Alterations of telomere length and DNA methylation in hairdressers: a cross-sectional study. Environ Mol Mutagen. 2016;57(2):159–67. doi:10.1002/em.21991.

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, et al. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell. 2008;133(3):523–36.

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009;462(7271):315–22. doi:10.1038/nature08514. nature08514 [pii].

Lister R, Pelizzola M, Kida YS, Hawkins RD, Nery JR, Hon G, et al. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature. 2011;471(7336):68–73. doi:10.1038/nature09798.

Liu Z, Liu S, Xie Z, Blum W, Perrotti D, Paschka P, et al. Characterization of in vitro and in vivo hypomethylating effects of decitabine in acute myeloid leukemia by a rapid, specific and sensitive LC-MS/MS method. Nucleic Acids Res. 2007;35(5):e31.

Lo YM, Wong IH, Zhang J, Tein MS, Ng MH, Hjelm NM. Quantitative analysis of aberrant p16 methylation using real-time quantitative methylation-specific polymerase chain reaction. Cancer Res. 1999;59(16):3899–903.

Lorthongpanich C, Cheow LF, Balu S, Quake SR, Knowles BB, Burkholder WF, et al. Single-cell DNA-methylation analysis reveals epigenetic chimerism in preimplantation embryos. Science. 2013;341(6150):1110–2. doi:10.1126/science.1240617.

Lun FM, Chiu RW, Sun K, Leung TY, Jiang P, Chan KC, et al. Noninvasive prenatal methylomic analysis by genomewide bisulfite sequencing of maternal plasma DNA. Clin Chem. 2013;59(11):1583–94. doi:10.1373/clinchem.2013.212274.

Lutsik P, Feuerbach L, Arand J, Lengauer T, Walter J, Bock C. BiQ Analyzer HT: locus-specific analysis of DNA methylation by high-throughput bisulfite sequencing. Nucleic Acids Res. 2011;39(Web Server issue):W551–6. doi:10.1093/nar/gkr312.

Lyko F, Foret S, Kucharski R, Wolf S, Falckenhayn C, Maleszka R. The honey bee epigenomes: differential methylation of brain DNA in queens and workers. PLoS Biol. 2010;8(11):e1000506. doi:10.1371/journal.pbio.1000506.

Manrao EA, Derrington IM, Pavlenok M, Niederweis M, Gundlach JH. Nucleotide discrimination with DNA immobilized in the MspA nanopore. PLoS One. 2011;6(10):e25723. doi:10.1371/journal.pone.0025723.

Marques-Rocha JL, Milagro FI, Mansego ML, Mourao DM, Martinez JA, Bressan J. LINE-1 methylation is positively associated with healthier lifestyle but inversely related to body fat mass in healthy young individuals. Epigenetics. 2016;11(1):49–60. doi:10.1080/15592294.2015.1135286.

Martinez-Arguelles DB, Lee S, Papadopoulos V. In silico analysis identifies novel restriction enzyme combinations that expand reduced representation bisulfite sequencing CpG coverage. BMC Res Notes. 2014;7:534. doi:10.1186/1756-0500-7-534.

Martino D, Joo JE, Sexton-Oates A, Dang T, Allen K, Saffery R, et al. Epigenome-wide association study reveals longitudinally stable DNA methylation differences in CD4+ T cells from children with IgE-mediated food allergy. Epigenetics. 2014;9(7):998–1006. doi:10.4161/epi.28945.

Masser DR, Berg AS, Freeman WM. Focused, high accuracy 5-methylcytosine quantitation with base resolution by benchtop next-generation sequencing. Epigenetics Chromatin. 2013;6(1):33. doi:10.1186/1756-8935-6-33.

Matarese F, Carrillo-de Santa Pau E, Stunnenberg HG. 5-Hydroxymethylcytosine: a new kid on the epigenetic block? Mol Syst Biol. 2011;7:562. doi:10.1038/msb.2011.95.

Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature. 2010;466(7303):253–7. doi:10.1038/nature09165.

Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res. 2005;33(18):5868–77. doi:10.1093/nar/gki901.

Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature. 2008;454(7205):766–70.

Melnikov AA, Gartenhaus RB, Levenson AS, Motchoulskaia NA, Levenson Chernokhvostov VV. MSRE-PCR for analysis of gene-specific DNA methylation. Nucleic Acids Res. 2005;33(10):e93.

Miceli-Richard C, Wang-Renault SF, Boudaoud S, Busato F, Lallemand C, Bethune K, et al. Overlap between differentially methylated DNA regions in blood B lymphocytes and genetic at-risk loci in primary Sjogren's syndrome. Ann Rheum Dis. 2016;75(5):933–40. doi:10.1136/annrheumdis-2014-206998.

Mikeska T, Candiloro IL, Dobrovic A. The implications of heterogeneous DNA methylation for the accurate quantification of methylation. Epigenomics. 2010;2(4):561–73. doi:10.2217/epi.10.32.

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 2007;448(7153):553–60.

Miller OJ, Schnedl W, Allen J, Erlanger BF. 5-Methylcytosine localised in mammalian constitutive heterochromatin. Nature. 1974;251(5476):636–7.

Miner BE, Stoger RJ, Burden AF, Laird CD, Hansen RS. Molecular barcodes detect redundancy and contamination in hairpin-bisulfite PCR. Nucleic Acids Res. 2004;32(17):e135. doi:10.1093/nar/gnh132.

Miura F, Ito T. Highly sensitive targeted methylome sequencing by post-bisulfite adaptor tagging. DNA Res. 2015;22(1):13–8. doi:10.1093/dnares/dsu034.

Miura F, Enomoto Y, Dairiki R, Ito T. Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. Nucleic Acids Res. 2012;40(17):e136. doi:10.1093/nar/gks454.

Moison C, Senamaud-Beaufort C, Fourriere L, Champion C, Ceccaldi A, Lacomme S, et al. DNA methylation associated with polycomb repression in retinoic acid receptor beta silencing. FASEB J. 2013;27(4):1468–78. doi:10.1096/fj.12-210971.

Moison C, Assemat F, Daunay A, Tost J, Guieysse-Peugeot AL, Arimondo PB. Synergistic chromatin repression of the tumor suppressor gene RARB in human prostate cancers. Epigenetics. 2014;9(4):477–82. doi:10.4161/epi.27869.

Moran S, Vizoso M, Martinez-Cardus A, Gomez A, Matias-Guiu X, Chiavenna SM, et al. Validation of DNA methylation profiling in formalin-fixed paraffin-embedded samples using the Infinium HumanMethylation450 Microarray. Epigenetics. 2014;9(6):829–33.

Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. Epigenomics. 2016;8(3):389–99. doi:10.2217/epi.15.114.

Morris TJ, Butcher LM, Feber A, Teschendorff AE, Chakravarthy AR, Wojdacz TK, et al. ChAMP: 450k chip analysis methylation pipeline. Bioinformatics. 2014;30(3):428–30. doi:10.1093/bioinformatics/btt684.

Mund C, Hackanson B, Stresemann C, Lubbert M, Lyko F. Characterization of DNA demethylation effects induced by 5-Aza-2′-deoxycytidine in patients with myelodysplastic syndrome. Cancer Res. 2005;65(16):7086–90.

Murphy PJ, Cipriany BR, Wallin CB, Ju CY, Szeto K, Hagarman JA, et al. Single-molecule analysis of combinatorial epigenomic states in normal and tumor cells. Proc Natl Acad Sci U S A. 2013;110(19):7772–7. doi:10.1073/pnas.1218495110.

Nair SS, Coolen MW, Stirzaker C, Song JZ, Statham AL, Strbenac D, et al. Comparison of methyl-DNA immunoprecipitation (MeDIP) and methyl-CpG binding domain (MBD) protein capture for genome-wide DNA methylation analysis reveal CpG sequence coverage bias. Epigenetics. 2011;6(1):34–44. doi:10.4161/epi.6.1.13313.

Nazor KL, Boland MJ, Bibikova M, Klotzle B, Yu M, Glenn-Pratola VL, et al. Application of a low cost array-based technique – TAB-Array – for quantifying and mapping both 5mC and 5hmC at single base resolution in human pluripotent stem cells. Genomics. 2014;104(5):358–67. doi:10.1016/j.ygeno.2014.08.014.

Nestor C, Ruzov A, Meehan R, Dunican D. Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA. Biotechniques. 2010;48(4):317–9. doi:10.2144/000113403.

Newman AM, Bratman SV, To J, Wynne JF, Eclov NC, Modlin LA, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nat Med. 2014;20(5):548–54. doi:10.1038/nm.3519.

Newton K, Jorgensen NM, Wallace AJ, Buchanan DD, Lalloo F, McMahon RF, et al. Tumour MLH1 promoter region methylation testing is an effective prescreen for Lynch Syndrome (HNPCC). J Med Genet. 2014;51(12):789–96. doi:10.1136/jmedgenet-2014-102552.

Nikolaidis G, Raji OY, Markopoulou S, Gosney JR, Bryan J, Warburton C, et al. DNA methylation biomarkers offer improved diagnostic efficiency in lung cancer. Cancer Res. 2012;72(22):5692–701. doi:10.1158/0008-5472.CAN-12-2309.

Nones K, Waddell N, Song S, Patch AM, Miller D, Johns A, et al. Genome-wide DNA methylation patterns in pancreatic ductal adenocarcinoma reveal epigenetic deregulation of SLIT-ROBO, ITGA2 and MET signaling. Int J Cancer. 2014;135(5):1110–8. doi:10.1002/ijc.28765.

Nygren AO, Ameziane N, Duarte HM, Vijzelaar RN, Waisfisz Q, Hess CJ, et al. Methylation-specific MLPA (MS-MLPA): simultaneous detection of CpG methylation and copy number changes of up to 40 sequences. Nucleic Acids Res. 2005;33(14):e128.

Nygren AO, Dean J, Jensen TJ, Kruse S, Kwong W, van den Boom D, et al. Quantification of fetal DNA by use of methylation-based DNA discrimination. Clin Chem. 2010;56(10):1627–35. doi:10.1373/clinchem.2010.146290. clinchem.2010.146290 [pii].

Oakes CC, La Salle S, Robaire B, Trasler JM. Evaluation of a quantitative DNA methylation analysis technique using methylation-sensitive/dependent restriction enzymes and real-time PCR. Epigenetics. 2006;1(3):146–52.

Ogino S, Kawasaki T, Brahmandam M, Yan L, Cantor M, Namgyal C, et al. Sensitive sequencing method for KRAS mutation detection by Pyrosequencing. J Mol Diagn. 2005;7(3):413–21. doi:10.1016/S1525-1578(10)60571-5.

Ogino S, Kawasaki T, Brahmandam M, Cantor M, Kirkner GJ, Spiegelman D, et al. Precision and performance characteristics of bisulfite conversion and real-time PCR (MethyLight) for quantitative DNA methylation analysis. J Mol Diagn. 2006;8(2):209–17.

Olsen JA, Kenna LA, Spelios MG, Hessner MJ, Akirav EM. Circulating differentially methylated amylin DNA as a biomarker of beta-cell loss in type 1 diabetes. PLoS One. 2016;11(4):e0152662. doi:10.1371/journal.pone.0152662.

Paliwal A, Temkin AM, Kerkel K, Yale A, Yotova I, Drost N, et al. Comparative anatomy of chromosomal domains with imprinted and non-imprinted allele-specific DNA methylation. PLoS Genet. 2013;9(8):e1003622. doi:10.1371/journal.pgen.1003622.

Papageorgiou EA, Fiegler H, Rakyan V, Beck S, Hulten M, Lamnissou K, et al. Sites of differential DNA methylation between placenta and peripheral blood: molecular markers for non-invasive prenatal diagnosis of aneuploidies. Am J Pathol. 2009;174(5):1609–18. doi:10.2353/ajpath.2009.081038. ajpath.2009.081038 [pii].

Pardo CE, Darst RP, Nabilsi NH, Delmas AL, Kladde MP. Simultaneous single-molecule mapping of protein-DNA interactions and DNA methylation by MAPit. Curr Protoc Mol Biol. 2011;Chapter 21:Unit 21 2. doi:10.1002/0471142727.mb2122s95.

Pastor WA, Pape UJ, Huang Y, Henderson HR, Lister R, Ko M, et al. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature. 2011;473(7347):394–7. doi:10.1038/nature10102.

Paul DS, Guilhamon P, Karpathakis A, Butcher LM, Thirlwell C, Feber A, et al. Assessment of RainDrop BS-seq as a method for large-scale, targeted bisulfite sequencing. Epigenetics. 2014;9(5):678–84. doi:10.4161/epi.28041.

Peat JR, Dean W, Clark SJ, Krueger F, Smallwood SA, Ficz G, et al. Genome-wide bisulfite sequencing in zygotes identifies demethylation targets and maps the contribution of TET3 oxidation. Cell Rep. 2014;9(6):1990–2000. doi:10.1016/j.celrep.2014.11.034.

Pelizzola M, Koga Y, Urban AE, Krauthammer M, Weissman S, Halaban R, et al. MEDME: an experimental and analytical methodology for the estimation of DNA methylation levels based on microarray derived MeDIP-enrichment. Genome Res. 2008;18(10):1652–9.

Petterson A, Chung TH, Tan D, Sun X, Jia XY. RRHP: a tag-based approach for 5-hydroxymethylcytosine mapping at single-site resolution. Genome Biol. 2014;15(9):456. doi:10.1186/s13059-014-0456-5.

Pham TT, Yin J, Eid JS, Adams E, Lam R, Turner SW, et al. Single-locus enrichment without amplification for sequencing and direct detection of epigenetic modifications. Mol Genet Genomics. 2016. doi:10.1007/s00438-016-1167-2.

Pogribny I, Yi P, James SJ. A sensitive new method for rapid detection of abnormal methylation patterns in global DNA and within CpG islands. Biochem Biophys Res Commun. 1999;262(3):624–8.

Popp C, Dean W, Feng S, Cokus SJ, Andrews S, Pellegrini M, et al. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature. 2010;463(7284):1101–5. doi:10.1038/nature08829.

Priolo M, Sparago A, Mammi C, Cerrato F, Lagana C, Riccio A. MS-MLPA is a specific and sensitive technique for detecting all chromosome 11p15.5 imprinting defects of BWS and SRS in a single-tube experiment. Eur J Hum Genet. 2008;16(5):565–71.

Pröll J, Födermayr M, Wechselberger C, Pammer P, Sonnleitner M, Zach O, et al. Ultra-sensitive immunodetection of 5′methyl cytosine for DNA methylation analysis on oligonucleotide microarrays. DNA Res. 2006;13(1):37–42.

Qui Y, Yang Q, Sui F, Lu R, Dang S, Ji M, et al. A strategy for accurate quantification of 5-methylcytosine and 5-hydroxymethylcytosine at CpG sites within gene promoter. J Biomed Nanotechnol. 2015;11(6):1016–26.

Radpour R, Kohler C, Haghighi MM, Fan AX, Holzgreve W, Zhong XY. Methylation profiles of 22 candidate genes in breast cancer using high-throughput MALDI-TOF mass array. Oncogene. 2009;28(33):2969–78. doi:10.1038/onc.2009.149. doi:10.1373/clinchem.2010.146290. onc2009149 [pii].

Ragoussis J, Elvidge GP, Kaur K, Colella S. Matrix-assisted laser desorption/ionisation, time-of-flight mass spectrometry in genomics research. PLoS Genet. 2006;2(7):e100.

Rand K, Qu W, Ho T, Clark SJ, Molloy P. Conversion-specific detection of DNA methylation using real-time polymerase chain reaction (ConLight-MSP) to avoid false positives. Methods. 2002;27(2):114–20.

Rand KN, Ho T, Qu W, Mitchell SM, White R, Clark SJ, et al. Headloop suppression PCR and its application to selective amplification of methylated DNA sequences. Nucleic Acids Res. 2005;33(14):e127.

Rauch T, Pfeifer GP. Methylated-CpG island recovery assay: a new technique for the rapid detection of methylated-CpG islands in cancer. Lab Invest. 2005;85(9):1172–80.

Riebler A, Menigatti M, Song JZ, Statham AL, Stirzaker C, Mahmud N, et al. BayMeth: improved DNA methylation quantification for affinity capture sequencing data using a flexible Bayesian approach. Genome Biol. 2014;15(2):R35. doi:10.1186/gb-2014-15-2-r35.

Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518(7539):317–30. doi:10.1038/nature14248.

Robertson AB, Dahl JA, Vagbo CB, Tripathi P, Krokan HE, Klungland A. A novel method for the efficient and selective identification of 5-hydroxymethylcytosine in genomic DNA. Nucleic Acids Res. 2011;39(8):e55. doi:10.1093/nar/gkr051.

Robertson AB, Dahl JA, Ougland R, Klungland A. Pull-down of 5-hydroxymethylcytosine DNA using JBP1-coated magnetic beads. Nat Protoc. 2012;7(2):340–50. doi:10.1038/nprot.2011.443.

Robinson MD, Stirzaker C, Statham AL, Coolen MW, Song JZ, Nair SS, et al. Evaluation of affinity-based genome-wide DNA methylation data: effects of CpG density, amplification bias, and copy number variation. Genome Res. 2010;20(12):1719–29. doi:10.1101/gr.110601.110. gr.110601.110 [pii].

Roessler J, Ammerpohl O, Gutwein J, Hasemeier B, Anwar SL, Kreipe H, et al. Quantitative cross-validation and content analysis of the 450k DNA methylation array from Illumina. Inc BMC Res Notes. 2012;5:210. doi:10.1186/1756-0500-5-210.

Ronaghi M. Pyrosequencing sheds light on DNA sequencing. Genome Res. 2001;11(1):3–11.

Ronaghi M, Uhlen M, Nyren P. A sequencing method based on real-time pyrophosphate. Science. 1998;281(5375):363, 5.

Ronn T, Ling C. DNA methylation as a diagnostic and therapeutic target in the battle against Type 2 diabetes. Epigenomics. 2015;7(3):451–60. doi:10.2217/epi.15.7.

Ross P, Hall L, Haff LA. Quantitative approach to single-nucleotide polymorphism analysis using MALDI-TOF mass spectrometry. Biotechniques. 2000;29(3):620–9.

Ross JP, Shaw JM, Molloy PL. Identification of differentially methylated regions using streptavidin bisulfite ligand methylation enrichment (SuBLiME), a new method to enrich for methylated DNA prior to deep bisulfite genomic sequencing. Epigenetics. 2013;8(1):113–27. doi:10.4161/epi.23330.

Rothbart SB, Strahl BD. Interpreting the language of histone and DNA modifications. Biochim Biophys Acta. 2014;1839(8):627–43. doi:10.1016/j.bbagrm.2014.03.001.

Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature. 2011;475(7356):348–52. doi:10.1038/nature10242.

Rougier N, Bourc'his D, Gomes DM, Niveleau A, Plachot M, Paldi A, et al. Chromosome methylation patterns during mammalian preimplantation development. Genes Dev. 1998;12(14):2108–13.

Ruzov A, Tsenkina Y, Serio A, Dudnakova T, Fletcher J, Bai Y, et al. Lineage-specific distribution of high levels of genomic 5-hydroxymethylcytosine in mammalian development. Cell Res. 2011;21(9):1332–42. doi:10.1038/cr.2011.113.

Saliba AE, Westermann AJ, Gorski SA, Vogel J. Single-cell RNA-seq: advances and future challenges. Nucleic Acids Res. 2014;42(14):8845–60. doi:10.1093/nar/gku555.

Salvaing J, Li Y, Beaujean N, O'Neill C. Determinants of valid measurements of global changes in 5′-methylcytosine and 5′-hydroxymethylcytosine by immunolocalisation in the early embryo. Reprod Fertil Dev. 2015;27(5):755–64. doi:10.1071/RD14136.

Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, et al. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics. 2011;6(6):692–702. doi:16196 [pii].

Schmidl C, Rendeiro AF, Sheffield NC, Bock C. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. Nat Methods. 2015. doi:10.1038/nmeth.3542.

Schreiber J, Wescoe ZL, Abu-Shumays R, Vivian JT, Baatar B, Karplus K, et al. Error rates for nanopore discrimination among cytosine, methylcytosine, and hydroxymethylcytosine along individual DNA strands. Proc Natl Acad Sci U S A. 2013;110(47):18910–5. doi:10.1073/pnas.1310615110.

Schultz MD, He Y, Whitaker JW, Hariharan M, Mukamel EA, Leung D, et al. Human body epigenome maps reveal noncanonical DNA methylation variation. Nature. 2015;523(7559):212–6. doi:10.1038/nature14465.

Schwarzenbach H, Hoon DS, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients. Nat Rev Cancer. 2011;11(6):426–37. doi:10.1038/nrc3066. nrc3066 [pii].

Scott RH, Douglas J, Baskcomb L, Nygren AO, Birch JM, Cole TR, et al. Methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA) robustly detects and distinguishes 11p15 abnormalities associated with overgrowth and growth retardation. J Med Genet. 2008;45(2):106–13.

Selker EU, Tountas NA, Cross SH, Margolin BS, Murphy JG, Bird AP, et al. The methylated component of the Neurospora crassa genome. Nature. 2003;422(6934):893–7.

Sengenes J, Daunay A, Charles MA, Tost J. Quality control and single nucleotide resolution analysis of methylated DNA immunoprecipitation products. Anal Biochem. 2010;407(1):141–3. doi:10.1016/j.ab.2010.07.013. S0003-2697(10)00459-8 [pii].

Serandour AA, Avner S, Mahe EA, Madigou T, Guibert S, Weber M, et al. Single-CpG resolution mapping of 5-hydroxymethylcytosine by chemical labeling and exonuclease digestion identifies evolutionarily unconserved CpGs as TET targets. Genome Biol. 2016;17(1):56. doi:10.1186/s13059-016-0919-y.

Serizawa RR, Ralfkiaer U, Dahl C, Lam GW, Hansen AB, Steven K, et al. Custom-designed MLPA using multiple short synthetic probes: application to methylation analysis of five promoter CpG islands in tumor and urine specimens from patients with bladder cancer. J Mol Diagn. 2010;12(4):402–8. doi:10.2353/jmoldx.2010.090152.

Serre D, Lee BH, Ting AH. MBD-isolated genome sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. Nucleic Acids Res. 2010;38(2):391–9. doi:10.1093/nar/gkp992.

Shapiro R, DiFate V, Welcher M. Deamination of cytosine derivatives by bisulfite. Mechanism of the reaction. J Am Chem Soc. 1974;96(3):906–12.

Shen L, Guo Y, Chen X, Ahmed S, Issa JP. Optimizing annealing temperature overcomes bias in bisulfite PCR methylation analysis. Biotechniques. 2007;42(1):48, 50, 2 passim.

Shim J, Humphreys GI, Venkatesan BM, Munz JM, Zou X, Sathe C, et al. Detection and quantification of methylation in DNA using solid-state nanopores. Sci Rep. 2013;3:1389. doi:10.1038/srep01389.

Shim J, Kim Y, Humphreys GI, Nardulli AM, Kosari F, Vasmatzis G, et al. Nanopore-based assay for detection of methylation in double-stranded DNA fragments. ACS Nano. 2015;9(1):290–300. doi:10.1021/nn5045596.

Shirane K, Toh H, Kobayashi H, Miura F, Chiba H, Ito T, et al. Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases. PLoS Genet. 2013;9(4):e1003439. doi:10.1371/journal.pgen.1003439.

Singer-Sam J, LeBon JM, Tanguay RL, Riggs AD. A quantitative HpaII-PCR assay to measure methylation of DNA from a small number of cells. Nucleic Acids Res. 1990;18(3):687.

Smallwood SA, Lee HJ, Angermueller C, Krueger F, Saadeh H, Peat J, et al. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat Methods. 2014;11(8):817–20. doi:10.1038/nmeth.3035.

Song CX, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, et al. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat Biotechnol. 2011;29(1):68–72. doi:10.1038/nbt.1732. nbt.1732 [pii].

Song CX, Clark TA, Lu XY, Kislyuk A, Dai Q, Turner SW, et al. Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine. Nat Methods. 2012;9(1):75–7. doi:10.1038/nmeth.1779. nmeth.1779 [pii].

Song CX, Szulwach KE, Dai Q, Fu Y, Mao SQ, Lin L, et al. Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. Cell. 2013;153(3):678–91. doi:10.1016/j.cell.2013.04.001.

Srinivasan M, Sedmak D, Jewell S. Effect of fixatives and tissue processing on the content and integrity of nucleic acids. Am J Pathol. 2002;161(6):1961–71. doi:10.1016/S0002-9440(10)64472-0.

Statham AL, Robinson MD, Song JZ, Coolen MW, Stirzaker C, Clark SJ. Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. Genome Res. 2012;22(6):1120–7. doi:10.1101/gr.132076.111.

Stefansson OA, Moran S, Gomez A, Sayols S, Arribas-Jorba C, Sandoval J, et al. A DNA methylation-based definition of biologically distinct breast cancer subtypes. Mol Oncol. 2015;9(3):555–68. doi:10.1016/j.molonc.2014.10.012.

Stelzer Y, Shivalila CS, Soldner F, Markoulaki S, Jaenisch R. Tracing dynamic changes of DNA methylation at single-cell resolution. Cell. 2015;163(1):218–29. doi:10.1016/j.cell.2015.08.046.

Stewart FJ, Panne D, Bickle TA, Raleigh EA. Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. J Mol Biol. 2000;298(4):611–22.

Stewart SK, Morris TJ, Guilhamon P, Bulstrode H, Bachman M, Balasubramanian S, et al. oxBS-450K: a method for analysing hydroxymethylation using 450K BeadChips. Methods. 2015;72:9–15. doi:10.1016/j.ymeth.2014.08.009.

Struhl K, Segal E. Determinants of nucleosome positioning. Nat Struct Mol Biol. 2013;20(3):267–73. doi:10.1038/nsmb.2506.

Sun Z, Terragni J, Borgaro JG, Liu Y, Yu L, Guan S, et al. High-resolution enzymatic mapping of genomic 5-hydroxymethylcytosine in mouse embryonic stem cells. Cell Rep. 2013;3(2):567–76. doi:10.1016/j.celrep.2013.01.001.

Sun Z, Dai N, Borgaro JG, Quimby A, Sun D, Correa Jr IR, et al. A sensitive approach to map genome-wide 5-hydroxymethylcytosine and 5-formylcytosine at single-base resolution. Mol Cell. 2015;57(4):750–61. doi:10.1016/j.molcel.2014.12.035.

Sun X, Chung TH, Tan D, Kim A. Practical guidelines and consideration of using RRHP for 5hmC detection. Epigenomics. 2016;8(2):225–35. doi:10.2217/epi.15.105.

Suzuki M, Jing Q, Lia D, Pascual M, McLellan A, Greally JM. Optimized design and data analysis of tag-based cytosine methylation assays. Genome Biol. 2010;11(4):R36. doi:10.1186/gb-2010-11-4-r36.

Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. Nucleic Acids Res. 2010;38(19):e181. doi:10.1093/nar/gkq684.

Taberlay PC, Kelly TK, Liu CC, You JS, De Carvalho DD, Miranda TB, et al. Polycomb-repressed genes have permissive enhancers that initiate reprogramming. Cell. 2011;147(6):1283–94. doi:10.1016/j.cell.2011.10.040.

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2009;324(5929):930–5. doi:10.1126/science.1170116. 1170116 [pii].

Taiwo O, Wilson GA, Morris T, Seisenberger S, Reik W, Pearce D, et al. Methylome analysis using MeDIP-seq with low DNA concentrations. Nat Protoc. 2012;7(4):617–36. doi:10.1038/nprot.2012.012.

Taiwo O, Wilson GA, Emmett W, Morris T, Bonnet D, Schuster E, et al. DNA methylation analysis of murine hematopoietic side population cells during aging. Epigenetics. 2013;8(10):1114–22. doi:10.4161/epi.26017.

Tangkijvanich P, Hourpai N, Rattanatanyong P, Wisedopas N, Mahachai V, Mutirangura A. Serum LINE-1 hypomethylation as a potential prognostic marker for hepatocellular carcinoma. Clin Chim Acta. 2007;379(1-2):127–33. doi:10.1016/j.cca.2006.12.029.

Taylor KH, Kramer RS, Davis JW, Guo J, Duff DJ, Xu D, et al. Ultradeep bisulfite sequencing analysis of DNA methylation patterns in multiple gene promoters by 454 sequencing. Cancer Res. 2007;67(18):8511–8.

Thirlwell C, Eymard M, Feber A, Teschendorff A, Pearce K, Lechner M, et al. Genome-wide DNA methylation analysis of archival formalin-fixed paraffin-embedded tissue using the Illumina Infinium HumanMethylation27 BeadChip. Methods. 2010;52(3):248–54. doi:10.1016/j.ymeth.2010.04.012.

Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, Webb S, et al. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature. 2010;464(7291):1082–6. doi:10.1038/nature08924.

Thomson JP, Hunter JM, Nestor CE, Dunican DS, Terranova R, Moggs JG, et al. Comparative analysis of affinity-based 5-hydroxymethylation enrichment techniques. Nucleic Acids Res. 2013;41(22):e206. doi:10.1093/nar/gkt1080.

Tobi EW, Goeman JJ, Monajemi R, Gu H, Putter H, Zhang Y, et al. DNA methylation signatures link prenatal famine exposure to growth and metabolism. Nat Commun. 2014;5:5592. doi:10.1038/ncomms6592.

Tost J. Epigenetics. Norwich: Horizon Scientific Press; 2008a.

Tost J. Methods for the genome-wide and gene-specific analysis of DNA methylation levels and patterns. In: Tost J, editor. Epigenetics. Norwich: Horizon Scientific Press; 2008b. p. 63–104.

Tost J, Gut IG. DNA methylation analysis by pyrosequencing. Nat Protoc. 2007;2(9):2265–75.

Tost J, Dunker J, Gut IG. Analysis and quantification of multiple methylation variable positions in CpG islands by Pyrosequencing. Biotechniques. 2003a;35(1):152–6.

Tost J, Schatz P, Schuster M, Berlin K, Gut IG. Analysis and accurate quantification of CpG methylation by MALDI mass spectrometry. Nucleic Acids Res. 2003b;31(9):e50.

Tost J, El abdalaoui H, Gut IG. Serial pyrosequencing for quantitative DNA methylation analysis. Biotechniques. 2006;40(6):721–2, 4, 6. doi:000112190 [pii].

Touleimat N, Tost J. Complete pipeline for Infinium((R)) human methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. Epigenomics. 2012;4(3):325–41. doi:10.2217/epi.12.21.

Tsui NB, Chiu RW, Ding C, El-Sheikhah A, Leung TN, Lau TK, et al. Detection of trisomy 21 by quantitative mass spectrometric analysis of single-nucleotide polymorphisms. Clin Chem. 2005;51(12):2358–62. doi:10.1373/clinchem.2005.056978. 51/12/2358 [pii].

Urich MA, Nery JR, Lister R, Schmitz RJ, Ecker JR. MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. Nat Protoc. 2015;10(3):475–83. doi:10.1038/nprot.2014.114.

Vaca-Paniagua F, Oliver J, Nogueira da Costa A, Merle P, McKay J, Herceg Z, et al. Targeted deep DNA methylation analysis of circulating cell-free DNA in plasma using massively parallel semiconductor sequencing. Epigenomics. 2015;7(3):353–62. doi:10.2217/epi.14.94.

Van Loo P, Nordgard SH, Lingjaerde OC, Russnes HG, Rye IH, Sun W, et al. Allele-specific copy number analysis of tumors. Proc Natl Acad Sci U S A. 2010;107(39):16910–5. doi:10.1073/pnas.1009843107.

Varinli H, Statham AL, Clark SJ, Molloy PL, Ross JP. COBRA-Seq: sensitive and quantitative methylome profiling. Genes (Basel). 2015;6(4):1140–63. doi:10.3390/genes6041140.

Varley KE, Mitra RD. Bisulfite Patch PCR enables multiplexed sequencing of promoter methylation across cancer samples. Genome Res. 2010;20(9):1279–87. doi:10.1101/gr.101212.109.

Vinarskaja A, Goering W, Ingenwerth M, Schulz WA. ID4 is frequently downregulated and partially hypermethylated in prostate cancer. World J Urol. 2012;30(3):319–25. doi:10.1007/s00345-011-0750-8.

Voelter-Mahlknecht S. Epigenetic associations in relation to cardiovascular prevention and therapeutics. Clini Epigenetics. 2016;8:4. doi:10.1186/s13148-016-0170-0.

Voss KH, Ross KP, Nonay NJ, Dovichi NJ. Combating PCR bias in bisulfite-based cytosine methylation analysis. Betaine-modified cytosine deamination PCt. Anal Chem. 1998;70:3818–23.

Vrielink A, Ruger W, Driessen HP, Freemont PS. Crystal structure of the DNA modifying enzyme beta-glucosyltransferase in the presence and absence of the substrate uridine diphosphoglucose. EMBO J. 1994;13(15):3413–22.

Walker DL, Bhagwate AV, Baheti S, Smalley RL, Hilker CA, Sun Z, et al. DNA methylation profiling: comparison of genome-wide sequencing methods and the infinium human methylation 450 bead chip. Epigenomics. 2015;7(8):1287–302. doi:10.2217/EPI.15.64.

Wallace EV, Stoddart D, Heron AJ, Mikhailova E, Maglia G, Donohoe TJ, et al. Identification of epigenetic DNA modifications with a protein nanopore. Chem Commun (Camb). 2010;46(43):8195–7. doi:10.1039/c0cc02864a.

Wang H, Guan S, Quimby A, Cohen-Karni D, Pradhan S, Wilson G, et al. Comparative characterization of the PvuRts1I family of restriction enzymes and their application in mapping genomic 5-hydroxymethylcytosine. Nucleic Acids Res. 2011a;39(21):9294–305. doi:10.1093/nar/gkr607.

Wang J, Jiang H, Ji G, Gao F, Wu M, Sun J, et al. High resolution profiling of human exon methylation by liquid hybridization capture-based bisulfite sequencing. BMC Genomics. 2011b;12:597. doi:10.1186/1471-2164-12-597.

Wang J, Xia Y, Li L, Gong D, Yao Y, Luo H, et al. Double restriction-enzyme digestion improves the coverage and accuracy of genome-wide CpG methylation profiling by reduced representation bisulfite sequencing. BMC Genomics. 2013a;14:11. doi:10.1186/1471-2164-14-11.

Wang Q, Gu L, Adey A, Radlwimmer B, Wang W, Hovestadt V, et al. Tagmentation-based whole-genome bisulfite sequencing. Nat Protoc. 2013b;8(10):2022–32. doi:10.1038/nprot.2013.118.

Wang J, Tang J, Lai M, Zhang H. 5-Hydroxymethylcytosine and disease. Mutat Res Rev Mutat Res. 2014;762:167–75. doi:10.1016/j.mrrev.2014.09.003.

Wang K, Li X, Dong S, Liang J, Mao F, Zeng C, et al. Q-RRBS: a quantitative reduced representation bisulfite sequencing method for single-cell methylome analyses. Epigenetics. 2015a;10(9):775–83. doi:10.1080/15592294.2015.107569.

Wang S, Lv J, Zhang L, Dou J, Sun Y, Li X et al. MethylRAD: a simple and scalable method for genome-wide DNA methylation profiling using methylation-dependent restriction enzymes. Open Biol. 2015b:5(11). doi:10.1098/rsob.150130.

Wanunu M, Cohen-Karni D, Johnson RR, Fields L, Benner J, Peterman N, et al. Discrimination of methylcytosine from hydroxymethylcytosine in DNA molecules. J Am Chem Soc. 2011;133(3):486–92. doi:10.1021/ja107836t.

Warnecke PM, Stirzaker C, Melki JR, Millar DS, Paul CL, Clark SJ. Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. Nucleic Acids Res. 1997;25(21):4422–6.

Warnecke PM, Stirzaker C, Song J, Grunau C, Melki JR, Clark SJ. Identification and resolution of artifacts in bisulfite sequencing. Methods. 2002;27(2):101–7.

Warren JD, Xiong W, Bunker AM, Vaughn CP, Furtado LV, Roberts WL, et al. Septin 9 methylated DNA is a sensitive and specific blood test for colorectal cancer. BMC Med. 2011;9:133. doi:10.1186/1741-7015-9-133.

Warton K, Samimi G. Methylation of cell-free circulating DNA in the diagnosis of cancer. Front Mol Biosci. 2015;2:13. doi:10.3389/fmolb.2015.00013.

Warton K, Lin V, Navin T, Armstrong NJ, Kaplan W, Ying K, et al. Methylation-capture and next-generation sequencing of free circulating DNA from human plasma. BMC Genomics. 2014;15:476. doi:10.1186/1471-2164-15-476.

Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet. 2005;37(8):853–62.

Weisenberger DJ, Trinh BN, Campan M, Sharma S, Long TI, Ananthnarayan S, et al. DNA methylation analysis by digital bisulfite genomic sequencing and digital MethyLight. Nucleic Acids Res. 2008;36(14):4689–98. doi:10.1093/nar/gkn455.

White HE, Hall VJ, Cross NC. Methylation-sensitive high-resolution melting-curve analysis of the SNRPN gene as a diagnostic screen for Prader-Willi and Angelman syndromes. Clin Chem. 2007;53(11):1960–2.

Wiebers JL. Detection and identification of minor nucleotides in intact deoxyribonucleic acids by mass spectrometry. Nucleic Acids Res. 1976;3(11):2959–70.

Wielscher M, Vierlinger K, Kegler U, Ziesche R, Gsur A, Weinhausel A. Diagnostic performance of plasma DNA methylation profiles in lung cancer, pulmonary fibrosis and COPD. EBioMed. 2015;2(8):927–34. doi:10.1016/j.ebiom.2015.06.025.

Wiencke JK, Bracci PM, Hsuang G, Zheng S, Hansen H, Wrensch MR, et al. A comparison of DNA methylation specific droplet digital PCR (ddPCR) and real time qPCR with flow cytometry in characterizing human T cells in peripheral blood. Epigenetics. 2014;9(10):1360–5. doi:10.4161/15592294.2014.967589.

Williams K, Christensen J, Pedersen MT, Johansen JV, Cloos PA, Rappsilber J, et al. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature. 2011;473(7347):343–8. doi:10.1038/nature10066.

Wojdacz TK, Dobrovic A. Methylation-sensitive high resolution melting (MS-HRM): a new approach for sensitive and high-throughput assessment of methylation. Nucleic Acids Res. 2007;35(6):e41.

Wojdacz TK, Hansen LL. Reversal of PCR bias for improved sensitivity of the DNA methylation melting curve assay. Biotechniques. 2006;41(3):274, 6, 8.

Wojdacz TK, Dobrovic A, Algar EM. Rapid detection of methylation change at H19 in human imprinting disorders using methylation-sensitive high-resolution melting. Hum Mutat. 2008a;29(10):1255–60. doi:10.1002/humu.20779.

Wojdacz TK, Dobrovic A, Hansen LL. Methylation-sensitive high-resolution melting. Nat Protoc. 2008b;3(12):1903–8. doi:10.1038/nprot.2008.191. nprot.2008.191 [pii].

Wojdacz TK, Borgbo T, Hansen LL. Primer design versus PCR bias in methylation independent PCR amplifications. Epigenetics. 2009;4(4):231–4.

Worm J, Aggerholm A, Guldberg P. In-tube DNA methylation profiling by fluorescence melting curve analysis. Clin Chem. 2001;47(7):1183–9.

Wu H, D'Alessio AC, Ito S, Wang Z, Cui K, Zhao K, et al. Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. Genes Dev. 2011;25(7):679–84. doi:10.1101/gad.2036011.

Xiao Y, Yu F, Pang L, Zhao H, Liu L, Zhang G, et al. MeSiC: a model-based method for estimating 5 mC Levels at single-CpG resolution from MeDIP-seq. Sci Rep. 2015;5:14699. doi:10.1038/srep14699.

Xiong Z, Laird PW. COBRA: a sensitive and quantitative DNA methylation assay. Nucleic Acids Res. 1997;25(12):2532–4.

Xu M, Kladde MP, Van Etten JL, Simpson RT. Cloning, characterization and expression of the gene coding for a cytosine-5-DNA methyltransferase recognizing GpC. Nucleic Acids Res. 1998;26(17):3961–6.

Yamada Y, Watanabe H, Miura F, Soejima H, Uchiyama M, Iwasaka T, et al. A comprehensive analysis of allelic methylation status of CpG islands on human chromosome 21q. Genome Res. 2004;14(2):247–66.

Yang AS, Estecio MR, Doshi K, Kondo Y, Tajara EH, Issa JP. A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. Nucleic Acids Res. 2004;32(3):e38.

Yang X, Dai W, Kwong DL, Szeto CY, Wong EH, Ng WT, et al. Epigenetic markers for noninvasive early detection of nasopharyngeal carcinoma by methylation-sensitive high resolution melting. Int J Cancer. 2015;136(4):E127–35. doi:10.1002/ijc.29192.

You JS, Kelly TK, De Carvalho DD, Taberlay PC, Liang G, Jones PA. OCT4 establishes and maintains nucleosome-depleted regions that provide additional layers of epigenetic regulation of its target genes. Proc Natl Acad Sci U S A. 2011;108(35):14497–502. doi:10.1073/pnas.1111309108.

Yu M, Hon GC, Szulwach KE, Song CX, Jin P, Ren B, et al. Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. Nat Protoc. 2012;7(12):2159–70. doi:10.1038/nprot.2012.137.

Yu L, Chibnik LB, Srivastava GP, Pochet N, Yang J, Xu J, et al. Association of brain DNA methylation in SORL1, ABCA7, HLA-DRB5, SLC24A4, and BIN1 with pathological diagnosis of Alzheimer disease. JAMA Neurol. 2015a;72(1):15–24. doi:10.1001/jamaneurol.2014.3049.

Yu M, Carter KT, Makar KW, Vickers K, Ulrich CM, Schoen RE, et al. MethyLight droplet digital PCR for detection and absolute quantification of infrequently methylated alleles. Epigenetics. 2015b;10(9):803–9. doi:10.1080/15592294.2015.1068490.

Yuan W, Xia Y, Bell CG, Yet I, Ferreira T, Ward KJ, et al. An integrated epigenomic analysis for type 2 diabetes susceptibility loci in monozygotic twins. Nat Commun. 2014;5:5719. doi:10.1038/ncomms6719.

Yuan Z, Claros CS, Suzuki M, Maggi EC, Kaner JD, Kinstlinger N, et al. Loss of MEN1 activates DNMT1 implicating DNA hypermethylation as a driver of MEN1 tumorigenesis. Oncotarget. 2016. doi:10.18632/oncotarget.7279.

Zeilinger S, Kuhnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. PLoS One. 2013;8(5):e63812. doi:10.1371/journal.pone.0063812.

Zeschnigk M, Bohringer S, Price EA, Onadim Z, Masshofer L, Lohmann DR. A novel real-time PCR assay for quantitative analysis of methylated alleles (QAMA): analysis of the retinoblastoma locus. Nucleic Acids Res. 2004;32(16):e125.

Zhang Y, Jeltsch A. The application of next generation sequencing in DNA methylation analysis. Genes (Basel). 2010;1(1):85–101. doi:10.3390/genes1010085.

Zhang Y, Rohde C, Tierling S, Jurkowski TP, Bock C, Santacruz D, et al. DNA methylation analysis of chromosome 21 gene promoters at single base pair and single allele resolution. PLoS Genet. 2009;5(3):e1000438. doi:10.1371/journal.pgen.1000438.

Zhang L, Szulwach KE, Hon GC, Song CX, Park B, Yu M, et al. Tet-mediated covalent labelling of 5-methylcytosine for its genome-wide detection and sequencing. Nat Commun. 2013;4:1517. doi:10.1038/ncomms2527.

Zhang Y, Yang R, Burwinkel B, Breitling LP, Brenner H. F2RL3 methylation as a biomarker of current and lifetime smoking exposures. Environ Health Perspect. 2014;122(2):131–7. doi:10.1289/ehp.1306937.

Ziller MJ, Gu H, Muller F, Donaghey J, Tsai LT, Kohlbacher O, et al. Charting a dynamic DNA methylation landscape of the human genome. Nature. 2013;500(7463):477–81. doi:10.1038/nature12433.

Zinn RL, Pruitt K, Eguchi S, Baylin SB, Herman JG. hTERT is expressed in cancer cell lines despite promoter DNA methylation by preservation of unmethylated DNA and active chromatin around the transcription start site. Cancer Res. 2007;67(1):194–201. doi:10.1158/0008-5472.CAN-06-3396.

DNA Methyltransferase Inhibitors: Development and Applications

Marie Lopez, Ludovic Halby, and Paola B. Arimondo

Abstract

As described in previous chapters of this book, DNA methylation is involved in numerous biological processes, and modulation of the activity of DNA methyltransferases (DNMTs) is a powerful strategy to modulate, restore, or reduce DNA methylation. In this chapter, we will present examples of inhibitors of DNMTs (DNMTi) and review the fields of applications of DNMTi mainly as therapeutic molecules, for example, in cancers, cardiovascular or neurological diseases, but also as bioengineering tools. Finally, the limits of currently available inhibitors will be discussed and the perspectives to discover improved DNMTi will be presented.

Abbreviations

5aza	5-azacytidine
5azadC	5-aza-2′-deoxycytidine
AD	Alzheimer's disease
ALI	Acute lung injury
ALL	Acute lymphoblastic leukemia
ALS	Amyotrophic lateral sclerosis
AML	Acute myeloid leukemia
Ara-C	Cytarabine
ASMA	Alpha-smooth muscle actin
ATRA	All-*trans* retinoic acid

M. Lopez • L. Halby • P.B. Arimondo (✉)
CNRS FRE3600 Epigenetic Targeting of Cancer ETaC – 3, avenue Hubert Curien – BP 13562, 31035 Toulouse Cedex 1, France
e-mail: paola.arimondo@etac.cnrs.fr

© Springer International Publishing Switzerland 2016

A. Jeltsch, R.Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*, Advances in Experimental Medicine and Biology 945,
DOI 10.1007/978-3-319-43624-1_16

BBB	Blood–brain barrier
BDNF	Brain-derived neurotrophic factor
BM	Bone marrow
CMML	Chronic myelomonocytic leukemia
CNS	Central nervous system
COMT	Catechol-*O*-methyltransferase
CpG	Cytosine–guanine dinucleotide
CVD	Cardiovascular disease
DNMT	DNA methyltransferase
DNMTi	DNMT inhibitor
EC	Endothelial cells
EGCG	(-)-epigallocatechin-3-gallate
EMA	European Medicines Agency
ER	Estrogen receptor
ESC	Embryonic stem cells
FDA	US Food and Drug Administration
GABA	γ-aminobutyric acid
GAD	Glutamic acid decarboxylase
HCV	Hepatitis C virus
HDAC	Histone deacetylase
HDACi	HDAC inhibitor
HIV	Human immunodeficiency virus
HPV	Human papilloma virus
HTS	High-throughput screening
INFγ	Interferon-γ
IVF	In vitro fertilization
LPS	Lipopolysaccharide
LTP	Long-term potentiation
MDS	Myelodysplastic syndrome
MeCP2	Methyl-CpG-binding protein 2
mPFC	Medial prefrontal cortex
MSC	Mesenchymal stromal cells
NHL	Non-Hodgkin's lymphomas
NMS	Neonatal maternal separation
NSC	Neuronal stem cell
PARP	Poly(ADP-ribose) polymerase
PARPi	PARP inhibitor
PD	Parkinson's disease
PTSD	Post-traumatic stress disorder
SAH	*S*-adenosyl-L-homocysteine
SAHA	Suberoylanilide hydroxamic acid, vorinostat (Zolinza®)
SAM	*S*-adenosyl-L-methionine
SAR	Seasonal allergic rhinitis
SFRP4	Secreted frizzled-related protein
SHS	Secondhand smoke

Th	T-helper
TSA	Trichostatin A
TSG	Tumor suppressor gene
VPA	Valproic acid
XRCC1	X-ray repair cross-complementing protein

1 How to Inhibit DNA Methyltransferases

To design inhibitors of DNMTs (DNMTi), it is important to analyze the mechanism of DNA methylation. Three actors are involved: the DNA, the enzyme and its cofactor, and the *S*-adenosyl-L-methionine (AdoMet). The DNMTs scan DNA and recognize CpG sites at which they flip out the deoxycytidine into the catalytic pocket. A cysteine (C1226 in human DNMT1, C711 in human DNMT3A, and C652 in human DNMT3B) binds then to position 6 of the cytosine (Fig. 1 top), followed by the transfer of the methyl group to position 5 of the cytosine (Fig. 2, gray arrow). Finally, by β-elimination the cysteine is released, and the enzyme is ready to start a new catalytic cycle (Fig. 1 top).

Accordingly, the catalytic pocket is composed of a cytidine and a cofactor binding site (Fig. 2). Mainly, DNMTs can be inhibited by targeting different parts of the catalytic pocket (the DNA, the AdoMet, or both) or by targeting an allosteric site.

When the DNA pocket is the target, for example, by DNA binders, it is important to design compounds specific for the DNMTs that do not inhibit other enzymes acting on DNA. A possibility is to use, for example, DNA binders specific for CpGs. An interesting alternative are compounds that recognize the cytidine binding site and, in addition, bear a chemical moiety able to react with the catalytic cysteine (cytidine analogs are the best example and are described in Sect. 2). Another

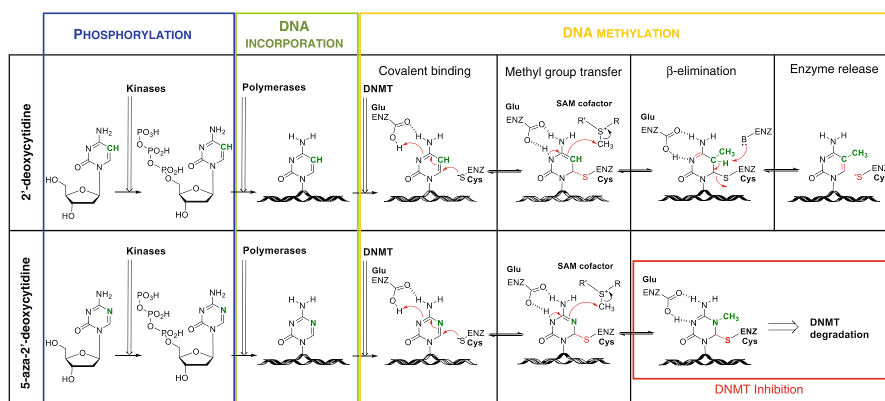


Fig. 1 Schematic representation of the catalytic cycles of DNMTs and mechanism of inhibition by 5azadC

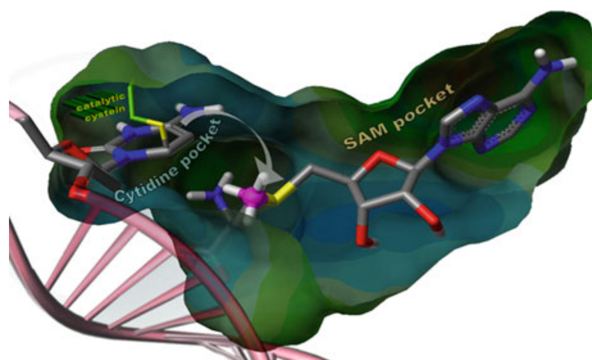


Fig. 2 Scheme of the transition state of DNA methylation (designed from the crystallographic structure of Dnmt1 available on PDB: 4DA4). In *red* is schematized the DNA duplex helix, and the catalytic cysteine of DNMT is shown in *green* and the SH group in *yellow*. The flipped-out deoxycytidine is shown in stick representation as the SAM. The methyl group that will be transferred (*gray arrow*) is represented in *pink*

strategy is to target the cofactor binding pocket by AdoMet analogs, for example. Since AdoMet is the most widely used enzyme cofactor after ATP (Schubert et al. 2003; Struck et al. 2012), a challenge consists in the design of compounds specific of the DNMT AdoMet-binding site and not binding to other methyltransferases. The design of AdoMet analogs that also bind to the deoxycytidine binding site could confer this specificity, and in addition it could potentially give higher affinity compounds and more potent inhibitors. Below we listed the principal inhibitors representing the different possible mechanisms of inhibition of DNMTs, at the exclusion of allosteric inhibitors since none have been described up to date.

2 Chemistry and Structure of DNMT Inhibitors

DNMTi are grouped in two families: the nucleoside inhibitors and the non-nucleoside ones (Fig. 3). To date, two compounds, *5-azacytidine* (*5aza*, Vidaza®) and *5-aza-2′-deoxycytidine* (*5azadC*, decitabine, Dacogen™), are approved by the FDA for the treatment of hematological cancers. 5Aza and 5azadC were the first DNMTi, and their story began with their synthesis in 1964 (Sorm et al. 1964). They were initially tested as antimetabolite agents against acute myelogenous leukemia (AML) (Sorm et al. 1964; Cihak 1974), with anticancer properties (Evans and Hanka 1968). However, it was in the late 1970s–early 1980s that a big turn was taken in the use of these compounds when Jones and Taylor showed that nontoxic doses 5aza and 5azadC induced cell differentiation by DNMT inhibition (Taylor and Jones 1979; Jones and Taylor 1980) (Fig. 4).

A new paradigm was found, as these compounds were shown to be able to reprogram cells, i.e., re-induce the expression of silenced genes and thus restore the normal functions of cells. Their mechanism of action was then understood identifying

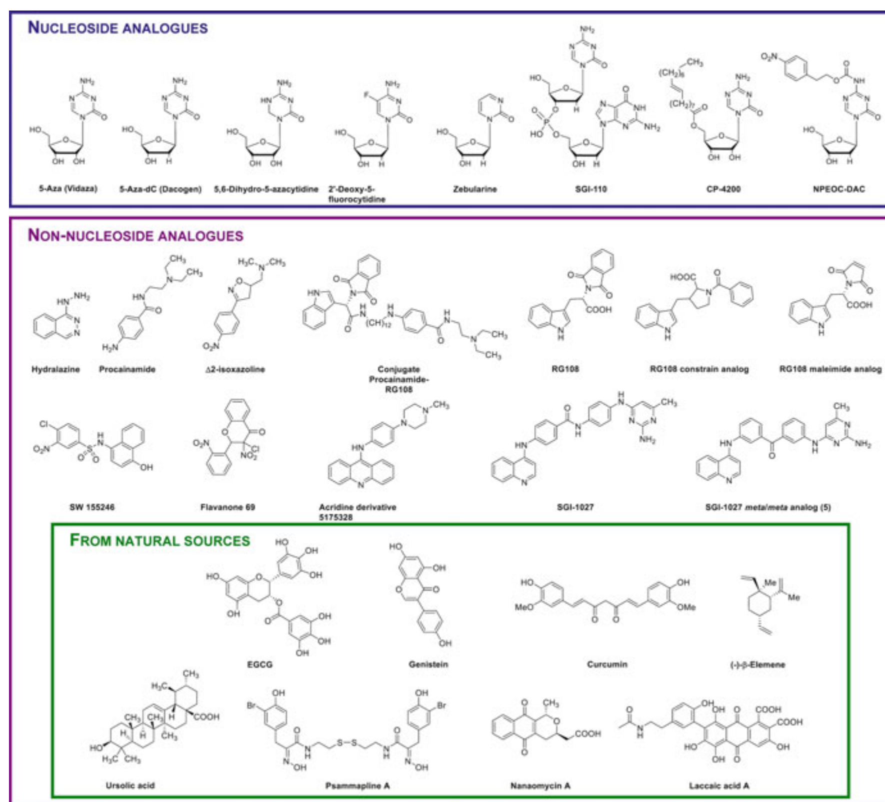


Fig. 3 Selection of compounds described as DNMT inhibitors

the aza-nucleosides as suicide substrates of DNMTs blocking the cytosine in the catalytic pocket of the enzyme (Fig. 2). More precisely, 5aza and 5azadC are triphosphorylated by kinases in cells (Momparler and Derse 1979) (Fig. 3), and then they are incorporated into DNA by the DNA polymerases (5aza is first modified to the deoxyribose) (Li et al. 1970). In DNA, at the CpG sites, 5azadC is recognized as deoxycytidine by the DNMTs and flipped out from the double helix into the catalytic pocket, and the catalytic cycle of DNMT occurs on this modified nucleoside. The catalytic cysteine of the DNMT binds covalently to the C-6 position, but, unlike to cytosine, the β-elimination, essential to release the enzyme, cannot occur after methyl group transfer from the AdoMet, because of the presence of the nitrogen at position 5. The enzyme is thus irreversibly trapped on the DNA and further degraded by the proteasome (Santi et al. 1984) inducing a demethylation of the DNA.

Based on these findings, 5aza (Vidaza®) and 5azadC (Dacogen™) were approved in 2004 and 2006, respectively, by the FDA for the treatment of acute myelodysplastic syndrome (MDS) and chronic myelomonocytic leukemia (CMML). However, these drugs are chemically and metabolically unstable, present a low bioavailability (Notari and Deyoung 1975; Chan et al. 1979), and as they are

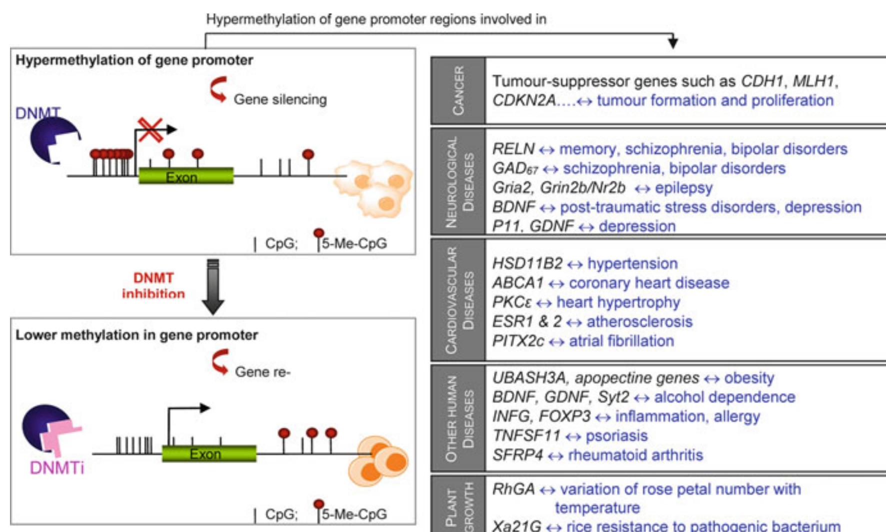


Fig. 4 Schematic representation of the mode of action of DNMTi as gene reexpressing agents and examples of DNA methylation-induced gene silencing in diverse pathologies. By inhibiting DNMTs, DNMTi can restore gene reexpression

incorporated at every deoxycytidine, they are not selective (Karahoca and Momparler 2013). Nevertheless, the proof of concept for DNMTs as a therapeutic target was established, and since then many efforts are dedicated to identify novel DNMTi, for which several strategies were adopted (see Sect. 1).

First, the chemical instability of 5aza and 5azadC was addressed by the design of more stable analogs, like *5,6-dihydro-5-azacytidine*, *2′-deoxy-5,6-dihydro-5-azacytidine*, *2′-deoxy-5-fluorocytidine*, and *zebularine*. *5,6-Dihydro-5-azacytidine* showed weaker inhibition of DNA methylation and was withdrawn from clinical trials (Yogelzang et al. 1997). Its analog *2′-deoxy-5,6-dihydro-5-azacytidine* was shown not to bind covalently to DNMTs, but simply to occupy the cytidine pocket (Sheikhnejad et al. 1999), resulting in the inhibition of methylation in several cell lines (CCRF-CEM and HL60) with very low cell toxicity (Matoušová et al. 2011). *2′-Deoxy-5-fluorocytidine* was described as more stable in aqueous media and as potent DNMTi in vitro, but it is currently rather recognized as a prodrug of a thymidylate synthase inhibitor, because it is metabolized into 5-fluorouridine (Jones and Taylor 1980; Boothman et al. 1989). *Zebularine* is also a stable compound with a weaker inhibition activity and cytotoxicity than azacytosine analogs (Cheng et al. 2003; Flotho et al. 2009). Its mode of action is different from 5aza and 5azadC, since it forms a very stable, but reversible complex with DNMTs that showed a slow dissociation kinetic (Champion et al. 2010; van Bemmel et al. 2009).

Second, to overcome the instability issue of 5azadC, a prodrug approach was chosen aiming at the release of the active molecule in the organism. This approach resulted in the generation of *NPEOC-DAC* (Byun et al. 2008), *CP-4200* (Brueckner et al. 2010), and *SG-110* (Yoo et al. 2007; Chuang et al. 2010). *SGI-110* or

guadecitabine, composed of a 5azadCpG dinucleotide, is a prodrug of 5azadC; it entered clinical trials and is the most promising compound among the prodrugs with a comparable in vitro and in vivo DNMT inhibition activity. The advantage of SGI-110 is that it is less sensitive to cytidine deaminase than 5azadC, which improves its metabolic stability.

Third, non-nucleoside compounds were investigated. Most of them were not promising because of their lack of specificity, their weak activity against DNMTs, or because they did not induce DNA methylation inhibition in cellular models. Several natural products were described as DNMTi (Fig. 3), for example, *(−)-epigallocatechin-3-gallate (EGCG)* (Fang et al. 2003), *nanaomycin A* (Kuck et al. 2010), and *laccaic acid* (Fagan et al. 2013). EGCG is a catechin extracted from green tea that showed DNA methylation inhibition activity, but its mode of action was questioned as the one of *genistein* (Fang et al. 2005). Indeed, Lee et al. demonstrated that these compounds did not act as direct catalytic DNMT inhibitors but as a substrate of AdoMet-dependent enzymes, catechol-*O*-methyltransferases (COMTs) (Lee et al. 2005b). The authors described that the metabolism of EGCG and genistein by COMTs induces an intracellular increase in the AdoHcy concentration responsible for the DNA methylation inhibition. Furthermore, two other research teams suggested that cytotoxicity induced by treatment of various cell lines with EGCG is simply caused by oxidative stress (Chuang et al. 2005; Stresemann et al. 2006). EGCG is a good example of a highly debated compound described as DNMTi. Other examples include genistein (Fang et al. 2005) and *curcumin* (Liu et al. 2009), because all these compounds are known as multi-target compounds and can induce DNA methylation inhibition by indirect effects. Moreover, Hann's team highlighted an indirect decrease in DNMT1 expression in cancer cells (lung cancer and hepatocarcinoma) via ERK1/ERK2- and AMPKa-mediated inhibition of the transcription factor Sp1 after treatment with two natural products, *β-elemene* (Zhao et al. 2015) and *ursolic acid* (Yie et al. 2015). *Psammaplin A* is also a nonspecific compound, and in-depth studies showed that it mainly acts as a histone deacetylase inhibitor (HDACi) and not DNMTi (Baud et al. 2012; Pereira et al. 2012). *Laccaic acid* is a highly substituted anthraquinone extracted from insects that inhibits DNMT1 and DNMT3A (Fagan et al. 2013). Noteworthy, while the compound induced some gene reexpression, no cellular DNA methylation inhibition was reported to accompany this gene reexpression. *Nanaomycin A* is a more specific compound, because it was found to target specifically DNMT3B and shown to induce genomic DNA methylation inhibition in colon cancer cell line (Kuck et al. 2010). Nevertheless, in a recent study, Penter et al. did not observe any DNA methylation inhibition by nanaomycin A in four neuroblastoma cell lines, although they observed an interesting synergistic activity when used in combination with doxorubicin (Penter et al. 2015). To conclude, none of the natural products described to date clearly targets DNMTs directly in cells.

The same difficulty to identify potent and specific non-nucleoside DNMTi is found among the synthetic compounds. For example, *hydralazine* (Cornacchia et al. 1988), *procainamide* (Cornacchia et al. 1988; Lee et al. 2005a), and *RG108* (Siedlecki et al. 2006; Stresemann et al. 2006; Brueckner et al. 2005) are weak

inhibitors of DNMTs (Chuang et al. 2005; Candelaria et al. 2012). Procainamide and hydralazine, FDA approved as vasodilator and anti-arrhythmic, respectively, were described as DNMTi because patients developed autoimmune disease (*lupus erythematosus*) correlated to an abnormal DNA methylation pattern of T-cells (Cornacchia et al. 1988). Even if hydralazine revealed a weak DNMT inhibition, a molecular modeling (Singh et al. 2009a) showed that hydralazine could be involved in a complex network of hydrogen bonds into the cytidine pocket. Today, hydralazine is still extensively studied, especially in combination with valproic acid (VPA), an HDAC inhibitor (Dueñas-Gonzalez et al. 2014). Procainamide was also described to potentially interact within the catalytic site of DNMTs by molecular modeling (Singh et al. 2009a), while biophysical studies previously demonstrated that it acted as DNA ligand and decreased the processivity of DNMTs (Lee et al. 2005a). Indeed procainamide and procaine were described as DNA ligand with a certain specificity for CG-rich regions, potentially targeting the DNA binding of the DNMTs (Villar-Garea et al. 2003). Procainamide was an interesting starting point for drug design, and several constrained derivatives were synthesized and tested (Castellano et al. 2008; Castellano et al. 2011). In parallel, procainamide conjugated to RG108 induced a synergy in the inhibition of the enzymatic activity of DNMT3A and DNMT1 (Halby et al. 2012). Initially it was hypothesized that the procainamide moiety of these compounds would bind to DNA and the RG108 moiety would bind to the DNMT catalytic pocket according to the discovery of RG108 by Brueckner et al. (Brueckner et al. 2005). Later a molecular modeling study suggested that the conjugates could occupy both AdoMet and cytidine pockets (Yoo et al. 2013) explaining the increased activity (cf. Sect. 1). Recent studies revealed that the promising RG108 has only a weak activity against human DNMTs (Halby et al. 2012; Suzuki et al. 2010), and more potent derivatives were since synthesized such as *maleimide* (Suzuki et al. 2010) and *constraint analogs* (Asgatay et al. 2014). Nevertheless, their ability to inhibit genomic DNA methylation and a potential correlation with tumor suppressor gene (TSG) reexpression has still to be established for all these second-generation inhibitors.

New-generation compounds seem to be more promising, as they showed DNA methylation inhibition and/or gene reexpression in cells. Several campaigns of high-throughput screening (HTS) were performed and allowed to identify new families of DNMTi. *3-Chloro-3-nitroflavanones* were identified (Ceccaldi et al. 2011) with sub-micromolar activity against Dnmt3a/3 L complex and exhibited a phenotype in zebrafish embryos similar to the one observed with the 5aza. A naphthoquinone, diclone, and other flavonoids were identified in another screening campaign, opening the path to study the impact on the epigenome of plants, animals, and humans when addressing the toxicology of pesticides (Ceccaldi et al. 2013). Molecular modeling studies suggested that the most potent compounds occupy both the cytosine and AdoMet pockets. *SW155246*, an aromatic sulfonamide, was also identified by HTS and showed a weak selectivity against DNMT1 vs. DNMT3A and DNMT3B (Kilgore et al. 2013). Interestingly, this compound was able to induce a weak methylation inhibition and reactivation of tumor suppressor genes (TSG) in human lung carcinoma. Acridine derivatives, known to intercalate into DNA, were

described to modulate DNA methylation (Hossain et al. 2013). Among them, derivative *517328* was the most potent analog, which inhibited DNMT1 in vitro and DNA methylation of TSGs in colon and pancreatic cancer cell lines. This compound provides a proof of concept for targeting DNA to inhibit DNMTs; however, the specificity remains an important issue for such inhibitors. *SGI-1027* is another weak DNA ligand that was identified to be a DNMT1 inhibitor (Datta et al. 2009). Molecular modeling studies suggested that SGI-1027 could inhibit DNMTs by occupying the cytidine and AdoMet pockets (Yoo et al. 2013), and Gros et al. (2015) confirmed by biophysical studies that it interacted with DNA and resulted in AdoMet noncompetitive but DNA-competitive inhibition. SGI-1027 was described to induce DNMT1 degradation and TSG reexpression in colon cancer cell lines and was quickly considered as the non nucleoside DNMTi starting point to expand syntheses of new derivatives (Gamage et al. 2013; Rilova et al. 2014; Valente et al. 2014). Valente et al. (2014) succeeded in increasing significantly the DNMT inhibition potency with the *meta/meta* analog, but no inhibition of genomic DNA methylation or TSG reexpression has so far been shown. Further investigations to characterize the mode of action of this compound family indicated that the *meta/meta* analog strongly interacted with DNA and inhibited DNMTs by DNA interaction and destabilization of the DNMT/DNA/AdoMet complex (Gros et al. 2015).

To conclude, to date only nucleoside DNMTi were FDA approved, and SGI-110 is a promising prodrug of 5azadC, evaluated in several clinical trials both on hematological and solid cancers. An increasing number of new non-nucleoside DNMTi was published in the last few years, but most lack an impact on genomic DNA methylation and TSG reexpression. Thus, more potent DNMTi need to be discovered.

3 Potential Applications of DNMT Inhibitors

DNMTs are responsible for DNA methylation, which is crucial for the control of gene expression and cell integrity. Indeed, when DNA methylation occurs on the CpG islands of gene promoters, the corresponding gene is silenced. This participates in the dynamics of the regulation of gene expression in cells. Therefore, DNMTs are crucial for normal cell functioning. However, many factors, including diet, stress, environmental conditions, etc., can lead to abnormal DNA methylation patterns. These deregulations are often responsible for dysfunctions and development of diseases such as cancers, nervous system and cardiovascular diseases, and abnormal plant growth (Fig. 3). As all epigenetic modifications, DNA methylation is reversible, and therefore modifications in the DNA methylation pattern can be corrected and reversed. Hence DNMTi can be powerful tools to restore a normal methylation profile and cell function.

The main application of DNMTi is by far their use to fight cancers. However, the importance of DNA methylation in various biological contexts leads to increasing research in other pathologies, such as neurodegenerative and cardiovascular diseases, and also in other fields, such as plant growth optimization.

3.1 DNMTi Application in Cancers

In cancer, a hypermethylation of TSG promoter regions was observed together with a global hypomethylation (Esteller 2008). The hypermethylation of the promoters, such as *P53*, *P16*, *P15*, *RARβ2*, *HIC1*, and *RASSF1A*, results in their silencing and participates to tumor formation, maintenance, and proliferation.

3.1.1 Nucleoside Analogs

As Single Agent

As described above (see Sect. 2), nucleoside analogs (Constantinides et al. 1977, 1978), namely, 5aza and 5azadC, are the most extensively used DNMTi and act as suicide substrates inducing DNMT degradation. A historical overview of their development and application is described in Issa and Kantarjian 2009. 5Aza and 5azadC are able to decrease the levels of TSG promoter methylation in cancer cells and induce their reexpression, resulting in cell reprogramming and, eventually, cell cycle arrest and apoptosis (Fahy et al. 2012). As single agents, 5aza and 5azadC are FDA and EMA approved to treat certain forms of leukemia and are in clinical trial for the treatment of solid tumors such as melanoma, breast, bladder, metastatic papillary thyroid, and follicular thyroid cancers (https://clinicaltrials.gov/ct2/home). Table 1 reports some examples of current clinical trials.

As described above, to overcome the drawbacks of 5aza and 5azadC (i.e., chemical instability, poor delivery, and side effects), chemically stable analogs were developed (Fig. 1), such as 5-fluoro-2′-deoxycytidine (Zhao et al. 2012) and zebularine (Savickiene et al. 2012), but they need to be used at higher doses due to their lower efficacy, and up to date no clinical trial was undertaken with these compounds (Yang et al. 2013). Nevertheless, the use of other analogs remains of interest, since it was shown that depending on the nucleoside analog used (5aza, 5azadC, or zebularine), the hypomethylation profile in the TSG promoters was different (Flotho et al. 2009).

Despite their promise for nonsolid tumor treatment, the direct use of nucleoside analogs as a single therapeutic agent is quite limited considering their instability, their side effects, and their lower therapeutic index in solid tumors.

Prodrugs of Nucleoside Analogs

Prodrugs of 5aza and 5azadC, which delay the release of the active molecule in the organism and lower its probability of degradation, showed a real improvement of the drugs. CP-4200 is a lipophilic ester of 5aza patented for its better cellular uptake (Silverman et al. 2009). In an in vivo orthotropic acute lymphoblastic leukemia (ALL) mouse model, a better therapeutic efficacy was observed compared to 5aza, but no clinical trial was undertaken up to date. SGI-110 (guadecitabine) is a 5azadC prodrug, which avoids its decomposition by deamination. It was already proven to be more stable and less toxic than 5azadC on tumor-free nude mice (Chuang et al. 2010) and is currently in several phase II clinical trials, phase III for AML (NCT02348489) and phase II for hepatocellular carcinoma (NCT01752933).

Table 1 Examples of anticancer clinical trials with DNMTi

	Status	Drugs	Type of cancer targeted (phase of study)
As a single agent	FDA approved	5Aza (Vidaza®)	AML, MDS
		5AzadC (Dacogen™)	AML, MDS, CMML
	In clinical trial	Guadecitabine	AML (III), hepatocellular carcinoma (II)
		5AzadC	NHL (I), CLL (I), solid tumors
		Disulfiram	Prostate cancer (I)
In combination	With other epidrug	5AzadC + SAHA	Hematological cancers (II)
		5AzadC + VPA	AML (II), MDS (II)
		5Aza + entinostat	Breast (II), metastatic colorectal cancers(II), lung (I)
	With chemotherapies	5Aza + Ara-C	AML (II), other hematological cancers
		5Aza + cisplatin	Ovarian cancer (II)
		5Aza + paclitaxel	Lung and neck cancers (I)
		CC-486 + abraxane (or gemcitabine)	Pancreatic cancer (I)
		CC-486 (oral 5aza) + carboplatin	Relapsed or refractory solid tumors (I)
	With immunotherapies	CC-486 + pembrolizumab (anti-PD1)	Non-small cell lung cancer (II)
		5AzadC + interferon2β	Renal cell carcinoma (II), metastatic solid tumors (I)
		Guadecitabine + ipilimumab (anti-CTLA-4)	Metastatic melanoma (I)

Phase of clinical trials are mentioned in brackets
AML acute myeloid leukemia, *MDS* myelodysplastic syndrome, *CLL* chronic lymphocytic leukemia, *NHL* non-Hodgkin's lymphoma

It is also used in combination with chemotherapy for ovarian cancer (NCT01696032) and with immunotherapy for metastatic melanoma treatment in Europe (EUdract 2015-001329-17).

Application of DNMTi in Combination with Other Drugs
Great promises arise from the use of epigenetic drugs in combination with chemotherapies or immune therapies in hematological as well as solid tumors (Table 1). The rational of these combination approaches relies on that fact that thanks to their capacity to affect globally the cells and restore cell functions, treatment with DNMTi increases sensitivity to other anticancer agents, increasing the treatment efficacy (Cameron et al. 1999; Azad et al. 2013; Ahuja et al. 2014, 2016).

Combination with Other Epidrugs
Combinations were studied with other epidrugs, as, for example, 5azadC in combination with trichostatin A (TSA), an HDACi, studied in colorectal carcinoma cell line. Whereas TSA was not able to cause reexpression of *MLH1*, *TIMP3*, *P15*, and *P16* genes, a pretreatment with a low dose of 5azadC led to their reexpression (Cameron et al. 1999). The combination of 5azadC and valproic acid (VPA), another HDACi, was evaluated in AML and MDS (Yang et al. 2005) and tested in phase II clinical trials (NCT00414310). But, in this case, the influence of VPA on 5azadC-treated MDS and AML patients did not improve the outcome. However, the association 5aza or 5azadC with the FDA-approved HDACi vorinostat (SAHA) was proven to be of interest in several ongoing phase I and phase II clinical trials in hematological cancers (Silverman et al. 2008). In vivo studies also exhibited a good synergy between 5azadC and SAHA in colon carcinoma cells (Yang et al. 2012). 5Aza was reported to be active in combination with entinostat and mocetinostat, benzamide inhibitors of HDACi (Fandy et al. 2009), and these combinations are tested in phase I and II clinical trials in hematological cancers. 5Aza/entinostat combination is also studied in solid cancers such as advanced non-small cell lung cancer (NCT01886573) (Juergens et al. 2011), advanced breast cancer (NCT01349959), and metastatic colorectal cancer (NCT01105377) (see Table 1 for some examples).

Combination with "Classical" Chemotherapies
A plethora of clinical trials involving multi-anticancer agents in combination with nucleoside DNMTi (5aza and 5azadC) are currently ongoing. For example, 5aza and 5azadC have been studied in combination with cytarabine (Ara-C). Ara-C/5azadC recently entered a phase II clinical trial for older AML patients (NCT01829503) (Table 1). 5Aza or 5azadC are also combined with all-*trans* retinoic acid (ATRA) (Xiang et al. 2014), but, like Ara-C, the combination is more frequently studied with additional therapeutic agents as, for example, the triple combination of 5aza or 5azadC/VPA/ATRA in phase II for AML. However, in this particular case, ATRA did not add any significant benefit to the 5aza or 5azadC/VPA treatment alone (Raffoux et al. 2010).

A phase I clinical trial is ongoing with CC-486 (oral 5aza) associated with carboplatin or ABI-007 (albumin-bound formulation of paclitaxel) for the treatment of relapsed or refractory solid tumors.

A low-dose, but long exposure to 5aza was described to sensitize chemoresistant cells to doxorubicine in diffuse large B-cell lymphoma (DLBCL) patients. The low-dose pretreatment results in cellular reprogramming, which increases sensitivity to doxorubicine without toxicity in vivo (Clozel et al. 2013). Therefore, CC-486 is also tested as a pretreatment to induce chemosensitivity. A phase II clinical trial is ongoing for the treatment of pancreatic cancer, where first-line chemotherapies (i.e., abraxane or gemcitabine) are administrated after 21–28 days of CC-486 treatment (NCT01845805).

By exploiting its gene reexpression potential, 5azadC is tested in clinical trial with plerixafor, an inhibitor of stromal cell-derived factor 1α (SDF-1α) in AML (phase I, NCT01352650), with panitumumab, an anti-EGFR, to overcome resistance to EGFR blocking agents in KRAS wild-type metastatic colorectal cancer (phase I, NCT00879385) or with interferon α-2b to stimulate immune response by reprogramming cells in metastatic solid tumors (phase I, NCT00701298), renal cell carcinoma (phase II, NCT00561912), or melanoma (phase I/II, NCT00791271). In triple negative metastatic breast cancer not responding to trastuzumab and tamoxifen, treatment with 5azadC/panobinostat (HDACi) restored the expression of estrogen factors and the efficiency of the tamoxifen treatment (phase I/II, NCT01194908). 5AzadC was also shown to sensitize human hepatoma cell line SMMC77221 to the cytotoxic effect of camptothecin (Ding et al. 2009).

Base excision repair mechanisms, involving X-ray repair cross-complementing protein 1 (XRCC1), could be able to excise 5azadC–DNMT adducts from the DNA by DNA repair pathways. Indeed, Orta et al. observed the co-localization of XRCC1 and DNMT1 in 5azadC-treated cells and an increase of 5azadC sensitivity in XRCC1-deficient cells (Orta et al. 2014). Furthermore, since poly(ADP-ribose) polymerase (PARP) inhibition prevents XRCC1 relocation to DNA damage sites, the authors associated 5azadC with a PARP inhibitor (olaparib). The combination of 5azadC/olaparib caused a synergetic lethality in leukemia cell lines (Orta et al. 2014). DNMTi and PARPi combinations are currently under investigation in other cancer models.

In parallel, several studies described an enhancement of radiosensitivity by 5azadC treatment in lung cancer A549 and glioblastoma U373MG cells (Kim et al. 2012), as well as SaOS2, HOS, and U2OS osteosarcoma cells (Li et al. 2014) and MDA-MB-231 and MDA-MB-435 breast cancer cell lines (Wang et al. 2013a). In all cases, this increase in radiosensitivity was correlated with gene upregulation and G2/M cell arrest.

In summary, nucleoside DNMTi are frequently used for hematological cancers. New clinical trials are ongoing to optimize the treatment schedule using a low-dose strategy and evaluate them in solid tumors. In particular, as a prodrug, SGI-110 could eventually overcome the compound stability limitation. In parallel, their use in combination with other anticancer agents is giving promising results in clinical trials and seems to increase the efficacy of the treatments and decrease the side effects, since lower doses of each drug are used.

3.1.2 Non-nucleoside DNMTi

As described above (see Sect. 2), several non-nucleoside DNMTi have been reported, but none obtained the same potency and validation as 5aza and 5azadC. Still, we believe that the in vivo data and the ongoing clinical trials with compounds that are not direct DNMTi but affect the DNA methylation profile of the cancer cells can be of interest for the better understanding of DNA methylation inhibition as therapeutic target.

This is the case, for example, with hydralazine (de la Cruz-Hernandez et al. 2011; Graça et al. 2014b). It failed as single therapeutic agent in phase II of breast and rectal cancer (Wang et al. 2009), but its combination with VPA (Duenas-Gonzalez et al. 2008) is in phase I against lung cancer (NCT00996060), in phase II for MDS (NCT01356875) (Candelaria et al. 2011), and in phase III against ovarian (NCT00533299) and cervical cancer (NCT00532818) (Song and Zhang 2009).

Disulfiram, previously known to be an aldehyde dehydrogenase inhibitor, was recently identified as a DNMTi in prostate cancer cell lines and in mice xenografts (Lin et al. 2011). In the presence of disulfiram, in vitro DNMT1 activity was shown to be reduced, and global hypomethylation was observed in prostate cancer cell lines, with decrease in *APC* and *RAR-β* promoter methylation accompanied by their reexpression. Xenograft tumor volumes were shown to be reduced under disulfiram treatment independently of the dose (from 10 to 40 $mg.kg^{-1}$). However, this study also reported a decrease in DNMT1 expression, which questioned the real target of disulfiram. Furthermore, DNMT inhibition properties of disulfiram are explored in an ongoing prostate cancer clinical trial (NCT01118741).

RG108, found by virtual screening, was shown to inhibit methylation in NALM6 (leukemia cell line) and in HCT116 (colorectal cancer cell line), inducing reexpression of *P16* and *TIMP3* (Stresemann et al. 2006). More recently, RG108 was tested in prostate cancer cell lines (LNCaP and 22Rv1), and a decrease of DNMT activity was observed together with cell growth inhibition and apoptosis (Graça et al. 2014a). However, as discussed above (see Sect. 2), RG108 is not a potent direct inhibitor of DNMTs, and the mechanism by which RG108 treatment resulted in a decrease of the methylation level in cells remains to be understood. Nevertheless, since it is commercially available, it was used as a tool to inhibit DNA methylation in several cellular models (see below).

Among natural products, curcumin, genistein, and EGCG are involved in a wide variety of clinical trials, but as described above they have multiple targets, and it is clear now that their effect on the DNA methylation patterns is indirect; therefore, their mechanism of action is controversial (Medina-Franco et al. 2011; Suh and Pezzuto 2012; Li and Tollefsbol 2010).

3.2 DNMTi Application in Neurological and Psychiatric Disorders

DNA methylation, together with other epigenetic modifications, is also deregulated in neurological diseases and psychiatric disorders, such as schizophrenia, epilepsy, or

bipolar disorder. However, to date, no DNMTi has been approved for these diseases, main hurdle being the poor capability of the existing DNMTi to cross the blood–brain barrier (BBB). Nevertheless, interesting results, listed below, highlight the great interest of targeting DNA methylation in neurological and psychiatric disorders.

3.2.1 Memory Formation

Memory loss is involved in many neurodegenerative diseases, such as Alzheimer's and Parkinson's diseases, schizophrenia, or post-traumatic stress disorders. DNA methylation is implicated in memory formation, as it regulates gene transcription in central nervous system (CNS) and is required in long-term memory formation (Day and Sweatt 2010). Additionally, DNA methylation was demonstrated to be involved in synaptic plasticity as *Dnmt1/Dnmt3a* double knockout in mice exhibit long-term potentiation (LTP) deficiencies (Feng et al. 2010). To better understand the implication of DNA methylation in various brain regions, rats were treated with RG108 by using the object-in-place paradigm, a test based on the memory of the positioning of objects in an enclosed space, which requires both hippocampus and perirhinal cortex brain regions. Using this model, Michnick et al. demonstrated that DNA methylation was required for long-term but not short-term memory (Mitchnick et al. 2015). In parallel, methylation of *Pp1*, a memory suppressor gene, and *Reln*, coding for reelin, a positive memory regulator, were shown to be controlled by promoter methylation. Following fear conditioning, an increase in *Pp1* promoter methylation and a decrease in *Reln* promoter methylation levels were observed in adult rat hippocampus. Infusion of 5aza in adult mouse brain inverted these effects, decreased *Pp1* methylation and enhanced a low methylation level of *Reln*, which inversely correlated with their expression (Miller and Sweatt 2007). In rats trained for contextual fear memory, the authors observed that intra-anterior cingulate cortex infusions of 5aza or zebularine, 30 days post-training, disrupted remote memory (Miller et al. 2010). Reelin is also implicated in synaptic plasticity that is involved in long-term storage memory. Levenson et al. demonstrated that zebularine induced a significant methylation decrease in one of the two CpG islands of the *Reln* promoter, whereas the other CpG methylation level remained steady (Levenson et al. 2006). The impact of DNMTi in synaptic plasticity was studied by Nelson et al. in hippocampus slices, where 5azadC was reported to decrease the genomic DNA methylation level, concomitantly with a diminution of miniature excitatory postsynaptic current frequencies in neurons, impacting neuronal activity (Nelson et al. 2008). Tetrodotoxin, known to decrease neuronal activity, also induced a decrease in methylation, and the same effects were observed both by using RG108 as DNMTi or by knocking down *Dnmt1* and *Dnmt3a* (Meadows et al. 2015).

Altogether, these studies show that DNA methylation is implicated in memory regulation, long-term memory, and synaptic plasticity. DNMTi were shown to interfere with the long-term memory and to enhance synapse receptiveness in neurons (synaptic upscaling). As a consequence, DNMTi were shown to impair memory, indicating that they could have a positive effect on post-traumatic stress disorders (PTSD) caused by long-lasting traumatic memory (see Sect. 3.2.5). Animal studies using 5azadC were, in most cases, carried out by direct injection in

brain regions or on slice cultures. However, 5azadC injected intraperitoneally in mice led also to an increase of *Bdnf* expression and a lower global DNA methylation in the hippocampus (Sales et al. 2011). These effects on memory functions were hypothesized as limiting factor for the use of decitabine in MDS patients (Aydin et al. 2012).

3.2.2 Schizophrenia

Schizophrenia is a cognition disorder often characterized by hallucination, paranoia, and failure to adopt a "normal" behavior in social situations. It is now quite clearly established that a deregulation of the (γ-aminobutyric acid) (GABA)ergic/glutamatergic network in the hippocampus and cortex is characteristic of a group of psychotic disorders including schizophrenia (Lewis et al. 2005). More specifically, a downregulation of GABAergic genes, such as *glutamic acid decarboxylase67* (*GAD67*) and *RELN*, was measured in postmortem samples of schizophrenia patients (Guidotti et al. 2000). This downregulation was correlated with a hypermethylation of their CpG island promoter regions (Chen et al. 2002; Grayson et al. 2005) and with an increase in DNMT1 (Veldic et al. 2004) and DNMT3A (Zhubi et al. 2009) mRNA expression in schizophrenia postmortem cortical GABAergic neurons. In a model of prenatal stress mice exhibiting a schizophrenia-like behavior, high levels of Dnmt1 and Dnmt3a expression occurred in GABAergic neurons (Matrisciano et al. 2013). Additionally, a knockdown of *Dnmt1* expression in mouse primary cortical cultures showed that *Reln* expression is controlled by Dnmt1 (Noh et al. 2005).

In this context, DNMTi were tested to decrease the methylation level and restore a normal expression of *GAD67* and *RELN*. First, 5aza, zebularine, and procainamide were tested in cell cultures, and an increase in the expression of *GAD67* and *RELN* and a decrease in *DNMT1* expression were observed (Kundakovic et al. 2007). However, 5aza and zebularine, the most active molecules in the cellular context, present limiting potential to cross the blood–brain barrier (BBB).

VPA, a well-known HDACi, was identified by Tremolizzo et al. as inducing a decrease in DNA methylation levels in mice with methionine-triggered schizophrenia alone (Tremolizzo et al. 2005) or in combination with clozapine, an antipsychotic agent (Guidotti et al. 2011). Noteworthy, in this case, the decrease in DNA methylation observed using an HDACi, which does not target DNMTs, occurs via an indirect pathway involving chromatin remodeling (Guidotti et al. 2009). Nicotine also induced a decrease in DNMT1 level and in *GAD67* promoter methylation in smoking schizophrenic patients (Satta et al. 2008).

3.2.3 Bipolar Disorders

Bipolar disorders are manic-depressive disorders characterized by a succession of elevated mood and depression periods. Aberrant DNA methylation patterns are implicated in the etiology of bipolar disorders, and, very similarly to schizophrenia, patients with bipolar disorders show a downregulation of *GAD67* and *RELN*. Because of the lack of relevant animal models to study bipolar disorders, very few

studies involving DNMTi have been carried out. Mainly, VPA was showed to decrease DNA methylation without decrease in DNMT expression (Aizawa 2015). VPA is FDA approved for bipolar disorders and in clinical studies in combination with antipsychotics.

3.2.4 Epilepsy

Epilepsy is a group of neurological diseases for which, in most cases, causes are unknown. Epileptic patients are subject to recurrent unprovoked seizures, which are brief to quite long periods of intense shaking.

In temporal lobe epilepsy, the most common epilepsy syndrome in human, low levels of reelin, a protein helping to maintain granule cells in the proper laminar structure expression, were observed in brain specimens (Heinrich et al. 2006) in correlation with a hypermethylation of its promoter region (Kobow et al. 2009). A genome-wide study revealed a global DNA hypermethylation in chronic epileptic rat correlated to a decrease in gene expression at the mRNA level (Kobow et al. 2013). *Gria2* encodes for a Glu2A subunit of ionotropic glutamate receptor, identified as playing a role in epilepsy, and its hypermethylation and corresponding decrease in mRNA expression was correlated to the intensity and frequency of seizure in rats (Machnes et al. 2013). In an equivalent rat model of kainic acid-induced epilepsy, bisulfite sequencing showed a hypermethylation of *Grin2b/Nr2b*, resulting in a lower level of the GRIN2B protein, another glutamate receptor subunit, together with a hypomethylation of the *Bdnf* (brain-derived neurotrophic factor) gene in epileptic hippocampus (Parrish et al. 2013). RG108 treatment on kainite-treated hippocampal slice cultures showed an inhibition of the *Gria2* hypermethylation and a beneficial increase in Gria2 activity (Machnes et al. 2013). Zebularine also decreased the methylation level of *Grin2b/Nr2b* concomitantly with an increase in GRIN2B protein level (Parrish et al. 2013). All these findings are in favor in applying DNMTi to epilepsy treatment.

3.2.5 Post-traumatic stress disorder

Post-traumatic stress disorders (PTSD) appear in patients who experienced a psychologically traumatic event, such as violent death witnessing, domestic violence, child abuse, etc., and result in severe anxiety associated with memory avoidance, flashbacks, nightmares, and emotional arousal. For example, war veterans commonly suffer from PTSD.

In PTSD patients, altered DNA methylation was observed in the peripheral blood immune cells. Together with other genes (Uddin et al. 2010), *BDNF* was reported to be hypermethylated (Smith et al. 2011). In rat models, epigenetic modifications were identified to play a role in fear memory, and Miller et al. demonstrated, in rat hippocampus, that *Dnmt3a* and *Dnmt3b* are upregulated, whereas level of Dnmt1 remains steady, in response to contextual fear conditioning (Miller and Sweatt 2007). Models of maltreated rats compared to normal animals also showed an increase in DNA methylation levels of *Bdnf*, associated with a decrease of *Bdnf* expression. This modification is a lasting effect, which was observed in adulthood

and passed to the next generation (Roth et al. 2009). DNA methylation of *BDNF* contributes to learned fear, and BDNF is important for the persistence of the pathological fear. Therefore, DNMTi could be envisaged to erase the memory of the trauma in PTSD patients (Zovkic and Sweatt 2013). Administration of DNMTi (5aza or zebularine) induces *Bdnf* expression even in adult animals. In PTSD, an over-consolidation of the fear memory is observed, and nucleoside DNMTi and RG108 were able to block the contextual fear memory formation. RG108 was also tested directly in rat brains and shown to abolish long-term fear memory (Miller et al. 2010).

In rodent, neonatal maternal separation (NMS) attenuates neuron differentiation and can lead to stress-induced behavioral disorders. This NMS results in a decrease in *RARα* expression concomitant with an increase in the *RARα* promoter methylation. *Dnmt1* was shown to be upregulated, whereas no change was measured in *Dnmt3a* and *Dnmt3b* expression. 5AzadC was capable of increasing the neuronal differentiation and decreasing methylation of *RARα* (Boku et al. 2015).

3.2.6 Depression

Sadness, low mood, loss of motivation, and low self-esteem characterize pathological depression, which, in the most severe cases, can lead to suicide. In the USA, 38,000 cases of suicide were reported in 2010, which represents more than the number of human beings killed by car accident (34,000) or twice more than victims of homicide (16,000) (Nature Editorials 2014). About 60 % of the subjects who committed suicide had depression or related mood disorders.

Until recently, the only epidrugs used to treat depression were HDACi (Sun et al. 2013). However, a couple of years ago, DNA methylation was proven to be involved in this disorder, and the environment was shown to play a major role, particularly early-life environmental stress (Booij et al. 2013). Dnmt3a was identified in a mouse model as regulating emotional behavior (LaPlant et al. 2010). In patients with severe depression and previous suicidal attempts, *BDNF* hypermethylation was proposed as a biological marker of suicidal behavior (Kang et al. 2013). Additionally, the gene encoding for P11 protein, a modulator of neuronal function involved in depression, was demonstrated to be hypermethylated in rodent and human depression. Interestingly, after treatment with antidepressant, such as escitalopram (a serotonin reuptake inhibitor), the methylation of *P11* was shown to return to normal (Melas et al. 2012). Hypermethylation of the promoter region of *TRKB*, encoding for tropomyosin receptor kinase B (TrkB), a BDNF receptor, was also reported in suicide completers accompanied by a decrease of mRNA expression (Ernst et al. 2009).

Sales et al. observed that 5aza, 5azadC, and RG108 possessed antidepressant-like effects on rats, as increasing the mobility in the tail suspension test and decreasing the immobility time in the forced swimming test (Sales et al. 2011). This was correlated to an increase of *BDNF* expression level. In a model of chronic ultra middle stress exposure in mice, representative of a depression-like behavior, zebularine and RG108 reverted the depression-like behavior with an increase in *Gdnf* expression (Uchida et al. 2011).

3.2.7 Autism Disorders

Patients with Rett syndrome, predominantly occurring in women, show a normal development for the first 6–18 months, then progressively loose speech and hand use, associated with seizure and autism. Rett syndrome is an X-chromosome-related neurological disease caused by a mutation in the methyl-CpG-binding protein 2 (MeCP2) protein (Amir et al. 1999; Guy et al. 2007). This protein is known to bind methylated DNA, to repress transcription, and to be involved in an activator model via the recruitment of the co-activator CREB1 (Lyst and Bird 2015). Although DNA methylation pattern is described as crucial for MeCP2 recognition, no DNMTi was used to treat Rett syndrome. However, treatment of hippocampal cells of MeCP2 knockout mice with methyl donor AdoMet was able to rescue neurotransmission event frequencies and thus to partially compensate MeCP2 loss of function (Nelson et al. 2008).

Additionally, in X-fragile syndrome, another autism disorder, a hypermethylation in the promoter region of the *FMR1* gene is observed (Sutcliffe et al. 1992). The treatment of X-fragile cell lines with 5azadC was reported to slightly re-induce *FRM1* expression (Tabolacci et al. 2005).

3.2.8 Parkinson's and Alzheimer's Diseases

Parkinson's disease (PD) is a CNS disease due to the degeneration of dopamine-producing neurons, and it affects more than four million people worldwide. The movement-related symptoms include shaking and dementia and can occur in a later stage of the disease. Hypomethylation is observed in PD patients in the CpG-rich island of *SNCA*, gene coding for α-synuclein (Jowaed et al. 2010). α-Synuclein is a protein forming aberrant soluble oligomers that lead to neuron death. This lower level of *SNCA* methylation was shown to result from a "sequestration" of DNMT1 in the cytosol in an α-synuclein transgenic mice and postmortem PD brains (Desplats et al. 2011). 5azadC was tested on dopaminergic neurons and resulted in viability decrease and increase of apoptosis associated with an upregulation of α-synuclein (Wang et al. 2013b). DNMTi can thus have deleterious effects for PD patients.

Alzheimer's disease (AD) is a neurodegenerative disorder resulting in severe dementia. The methylation profile of AD disease patients is controversial as both hypo- and hypermethylation were reported (Coppieters and Dragunow 2011). Nuclear immunostaining of DNMT1 and other component of the methylation machinery such as MBD2/MBD3 was significantly diminished in neurons (Mastroeni et al. 2010) where a loss of methylation of the amyloid precursor proteins was observed. A decrease in *BACE* and *PSEN1* (encoding for presenilin 1) methylation was also shown in AD patients and was associated with a potential overexpression of amyloid β-peptides (Scarpa et al. 2003; Mastroeni et al. 2010). A cocktail of DNA methylation enhancers, such as folate, is currently in phase III clinical trial (NCT00056225). Moreover, a dramatic global hypermethylation in the gray matter of postmortem human brain tissues was also observed in AD patients (Coppieters et al. 2014) together with an increase in methylation in certain regions such as *MCF2L* and *ANK1* genes (De Jager et al. 2014). Additional data are crucial

to confirm the hypermethylation profile of AD patient and to potentially envisage the use of DNMTi for AD treatment.

3.2.9 Aging-Related Senescence and Amyotrophic Lateral Sclerosis

In amyotrophic lateral sclerosis (ALS), motor neuron death leads to muscle weakness and evolves to breathing difficulties. This neurodegenerative disease, often associated with aging, has a 3–5-year survival prognostic (Cleveland and Rothstein 2001).

In motor cortex tissues and spinal cord motor neurons of ALS patients, levels of DNMT1 and DNMT3A were shown to be increased (Chestnut et al. 2011; Cho et al. 2010). Oh et al. found a decrease in cell potency in ALS mesenchymal stromal cells (ALS-MSCs) with a twice higher level of DNMT1 and DNMT3A compared to normal MSCs. RG108 treatment of ALS-MSCs isolated from ALS patients' bone morrow induced anti-senescence factors (*TERT*, *VEGF*, *ANG*), together with the downregulation of senescence factors (*ATM*, *P21*). Additionally, a significant improvement in ALS-MSC migration and their differentiation into neurons was observed (Oh et al. 2015b). Thus, RG108 treatment is promising for a more efficient ALS treatment by autologous cell therapies.

In addition, Oh et al. evaluated the effect of RG108 on human bone marrow mesenchymal stromal cells (hBM-MSCs), which are used in cell therapies. They confirmed the anti-senescence effect of RG108 in in vitro cultures with an increase in the expression of anti-senescence factors (*TERT*, *VEGF*, *bFGF*, and *ANG*) and a decrease of senescence-related factors (*ATM*, *P21*, and *P53*) (Oh et al. 2015a). Thus, an optimized dose of RG108 (5 μM), for which cell viability is maintained, could greatly improve hBM-MSC potency, which could constitute a real progress in the improvement of the stem cell therapies.

3.2.10 Neuronal Stem Cell

Neuronal stem cell (NSC) cultures are potential sources of transplantable cells to treat neurodegenerative diseases. DNA methylation was proven to be essential for NSC differentiation and proliferation. Folic acid was shown to stimulate neonatal rat NSC proliferation in vitro (Li et al. 2013), whereas zebularine treatment resulted in the attenuation of their proliferation (Lin et al. 2014; Luo et al. 2013) and a reduction of their migration (Singh et al. 2009b). Hence, DNMTi are deleterious for NSC growth, but the use of DNMT activators can provide an interesting way to optimize NSC cultures.

In summary, DNA methylation was shown to be implicated in many neurological diseases and psychiatric disorders. The consequences of DNA methylation depends on the disease, and the use of DNMTi can be, in a few cases, unfavorable, like in Alzheimer's disease, but it can have a positive outcome in pathologies like schizophrenia, depression, and post-traumatic stress disorders. Despite the high interest in controlling DNMT activity, the path to use DNMTi as drugs to treat neurological diseases and psychiatric disorders is still long. First, it is made difficult by the lack of animal models to study these pathologies. Second, active molecules must cross the blood–brain barrier, and, for example, 5aza or zebularine are not able to do so.

Additionally, toxicity is not acceptable for psychiatric diseases, which often require chronic and lifelong treatment. Despite these limitations, the better comprehension of DNA methylation in neurological diseases will lead to new therapeutic strategies, and new small molecule DNMTi are urgently needed for neurodegenerative diseases and psychiatric disorders (Szyf 2015).

3.3 DNMTi Application in Cardiovascular Diseases

Cardiovascular diseases (CVDs) are the leading cause of death in most of the developed countries and environmental factors such as diet, smoking habits, or impaired metabolism are critical risk factors. Therefore, it is not surprising that epigenetic modifications, including DNA methylation, are involved in these pathologies.

DNA methylation patterns were reported to be altered in several CVDs. In hypertension, the hypermethylation of the *HSD11B2* promoter, a gene affecting blood pressure, was observed (Friso et al. 2008). Hypermethylation was also reported for the ATP-binding cassette transporter A1 gene (*ABCA1*) in coronary heart disease (Guay et al. 2014), and in an atherosclerosis in *Apoe* knockout mouse model, the hypermethylation of specific vascular homeostasis genes was described (Zaina et al. 2014). A study on about 300 Singapore Chinese subjects also proposed DNA methylation as CVD risk biomarker, specifically the *Alu*/*STAT2* methylation level (Kim et al. 2010). In parallel, a global hypomethylation was reported in peripheral blood mononuclear cells (Lund et al. 2004). Concerning the use of DNMTi in CVDs, 6-day treatment with 5azadC restored a normal protein expression profile of the whole cardiac proteome and rescued the phenotype of norepinephrine-treated rats (Xiao et al. 2014).

Atherosclerosis is characterized by an inflammation of arterial walls, and it is the major cause of stroke and heart attack. Gene expression in endothelial cells changes dramatically when submitted to large blood flow variations, and disturbed flow is pro-atherogenic. In mice, disturbed blood flow induces *Dnmt* expression, and several downregulated mechanosensitive genes were identified, and their expression was shown to be under their promoter methylation control and reversed by 5azadC (Dunn et al. 2014). Besides, estrogen receptors (ERα, ERβ) are atheroprotective, and hypermethylation of *ESR1* (ERα) and *ESR2* (ERβ) promoters, associated with their silencing, was shown in atherosclerosic tissues and in senescing cells. Treatment with 5azadC was carried out on smooth muscle cells (SMCs) and endothelial cells (ECs), resulting in the decrease of estrogen receptor gene promoter methylation accompanied by an increase of ER levels in both vascular cell lines. Additionally, the combination of 5azadC/TSA showed some synergetic effect, while TSA alone had no effect (Kim et al. 2007).

Ischemia is a decrease in blood supply to tissue, which results in hypoxia (i.e., deficiency in oxygen supply) and leads to cardiac fibrosis. In hypoxia-induced pro-fibrotic states of human cardiac fibroblast, an increase in *DNMT1* and *DNMT3B* expression and a global hypermethylation were reported. The expression of pro-fibrotic genes, such as *alpha-smooth muscle actin* (*ASMA*) or *collagen 1*, increased,

and this was enhanced by treatment with TGFβ, a pro-fibrotic cytokine. Interestingly, treatment with 5azadC significantly reduced TGFβ effects and levels of ASMA and collagen 1, decreasing pro-fibrotic effects and positioning DNMT as potentially valuable therapeutic target in ischemic heart disease (Watson et al. 2014). Finally, exploring the influence of DNMTi on mouse embryonic stem cells (mESCs), zebularine was shown to reexpress cardiac-specific genes, such as *Nkx2.5* and *Gata4* (Horrillo et al. 2013).

DNA methylation is also implicated in heart failure. Indeed, a higher level of homeobox gene *PITX2c* promoter methylation with increased DNMT1 and decreased PITX2 protein levels was identified in heart disease atrium. In the HL-1 cell line, 5azadC reversed *PITX2c* promoter methylation and increased the PITX2 and KIR2.1 protein level (Kao et al. 2013), which could have a positive outcome in case of heart failure.

Despite the evidence of the role of DNA methylation in CVDs (Chaturvedi and Tyagi 2014), no epidrug is currently in clinical trial for CVDs (Schiano et al. 2015). Only the nonspecific multi-target polyphenols in cocoa, reported to decrease DNA methylation level in CVD patients (Crescenti et al. 2013), completed phase 3 clinical trial (NCT00511420) as a diet supplement.

3.4 DNMTi Application in Other Human Pathologies

3.4.1 Obesity

Obesity has a high prevalence in industrialized countries where high caloric diet is common. It constitutes a major public health problem and can lead to diabetes or cardiovascular diseases.

Following a 5-day high-fat diet, DNA methylation changes of the transcriptional co-activator PGC-1α involved in oxidative energy metabolism were observed (Brøns et al. 2010). Variations of DNA methylation were also reported by Wang et al., who analyzed DNA methylation in peripheral blood leukocytes and determined that *UBASH3A* was hypermethylated, whereas *TRIM3* was hypomethylated in obesity cases (Wang et al. 2010b). Variability of DNA methylation in individual adipose tissues was shown to influence their response to caloric restriction in terms of weight loss (Bouchard et al. 2010). A comparison of methylation levels in abdominal adipose tissues before and after gastric bypass and weight loss showed a decrease of DNA methylation in both tissues after gastric bypass (Benton et al. 2015). Expression of adiponectin, a protein regulating glucose and lipid metabolism (Yamauchi et al. 2002), was inversely correlated with insulin resistance, type 2 diabetes, and cardiovascular diseases (Kadowaki et al. 2006). Kim et al. established that adiponectin expression was under epigenetic control, by hypermethylation of the R2 promoter region, in high-fat diet obese mice compared to normal-diet lean mice. RG108 was tested in an obese mice model and increased adiponectin expression level leading to an improvement of glucose intolerance and insulin resistance (Kim et al. 2015). These recent results are promising for obesity-related disease therapeutics.

3.4.2 Alcohol Addiction

Alcohol abuse with its associated diseases and behaviors is responsible for about 6 % of the death worldwide (WHO 2014, Global status report on alcohol and health). In the UK, alcohol was classified as the most harmful drug for oneself and others (Nutt et al. 2010). The influence of alcohol consumption on DNA methylation is controversial and seems to be highly dependent on the studied tissues. For example, DNA in postmortem human brains was reported to be hypomethylated in alcoholics (Ponomarev et al. 2012). This hypomethylation was shown to result in the increased expression of long-term repeat-containing human endogenous retrovirus. A decrease in DNMT3A and DNMT3B level was also shown in alcoholic patients compared to healthy individuals. DNMT3B expression was reported to be inversely proportional to blood alcohol concentration, whereas no variation was reported for DNMT1 expression level (Bönsch et al. 2006). Genome-wide DNA hypermethylation in lymphoblast of 165 female subjects showed an increase of DNA methylation with the increase in drinking frequency (Philibert et al. 2012).

DNMTi were studied for the treatment of alcohol addiction and encouraging outcomes were described. In an alcohol-exposed murine model submitted to the two-bottle choice test and treated or not with 5aza, Warnault et al. found that for 5aza-treated animals, the rate of alcohol intake was significantly lower compared to non-treated alcohol-dependent rats (Warnault et al. 2013). These results also showed that inhibition of DNA methylation enabled to specifically reduce alcohol consumption and preference, but did not influence other rewarding substance intake. DNA methylation was suggested to increase the expression of endogenous factors interfering with alcohol drinking behavior, such as BDNF (Logrip et al. 2009) and GDNF (Carnicella et al. 2008) that prevent the escalation from moderate to excessive consumption. Barbier et al. studied long-term behavior on alcohol-conditioned rats after 3 weeks of abstinence and established a causal relationship between DNA methylation, alcohol intake, and seeking behavior. They measured an increase in Dnmt1 and no change in the Dnmt3A and Dnmt3B levels. Using intracerebroventricular infusion of RG108, they lowered the DNA methylation level in medial prefrontal cortex (mPFC) of alcohol-conditioned rat, and no change was observed in controls. DNMTi also prevented alcohol escalation. Seven genes involved in neurotransmission and coding for synaptic proteins were downregulated in alcohol-dependent rats. Among them, synaptotagmin 2 (*Syt2*) presented a higher methylation level in alcohol post-dependent rats, and a synaptic transmission deregulation was reported in post-dependent mPFC neurons. A *Syt2* knockdown resulted in compulsive-like drinking behavior. Interestingly, *Syt2* normal methylation level and expression were restored by RG108 treatment (Barbier et al. 2015).

Although variations of DNA methylation levels in the brain of alcoholic patients seem to be heterogeneous, the results of DNMTi on preclinical animal models are an interesting starting point for the use of DNMTi to limit alcohol dependence.

3.4.3 Inflammation and Allergy

Inflammation is a reaction of the immune system to harmful stimuli. In allergy, also known as hypersensitivity (e.g., hay fever, allergic asthma), inflammation is

triggered by an unsuitable immune response to environmental cues. Similarly, in autoimmune diseases, an abnormal immune response is directed toward tissues normally present in the organism. Among the $CD4^+$ T-cells (T-helper cells), T-helper 1 (Th1) express specifically interferon-γ (INFγ) (Brand et al. 2012), whereas T-helper 2 (Th2) express interleukin 4 (IL4), IL5, and IL13. The INFγ (Winders et al. 2004) and IL4 (Kwon et al. 2008) cytokines are under DNA methylation control, and differentiation of naïve T-cell into Th1 or Th2 cells results in a modification of the methylation level of CpG sites in *INFG*, the INFγ gene (White et al. 2006). The level of *INFG* promoter methylation was shown to depend on age, sex, and tissues (Lovinsky-Desir et al. 2014). In mouse models, *Dnmt3a* and *Dnmt3b* were more expressed in Th2 cells compared to Th1 cells, whereas no difference in *Dnmt1* expression was observed (Yu et al. 2012). Moreover, in the same study, *INFG* and *IL4* expression was lower in mutant mice lacking Dnmt3a or Dnmt3a/Dnmt3b but not in mutants lacking only Dnmt3b. Altogether these data highlight the importance of DNA methylation in the regulation of cytokine expression in $CD4^+$ T-cells and in inflammation reactions.

In inflammation, the balance of T-cell subtype is disturbed with a shift from Th1 to Th2 response. As a result the ratio INFγ vs. IL4 production is modified, which can be considered as an indicator of inflammation risk (Shahid et al. 2002). In patients with bronchial asthma, T-cell differentiation into the Th2 subtype and IL4 associated cytokine production are much higher than in controls (Kwon et al. 2008).

Although genetics are involved in asthma (Lloyd and Hawrylowicz 2009), the environment such as pollution, cigarette smoke, etc., is also well known to have a crucial influence, which rises the importance of epimutations in asthmatic patients. A comparison of 21 monozygotic adult twins discordant for asthma showed a decrease in *INFG* and *FOXP3* expression in the case of asthma with a greater decrease for cases with secondhand smoke (SHS) during childhood, which was shown to significantly affect the methylation levels (Runyon et al. 2012; Kohli et al. 2012). A similar study with monozygotic twins concluded the importance of DNA methylation of genes involved in immune response and associated with psoriasis, like *TNFSF11*, in $CD4^+$ T-cells (Gervin et al. 2012). It is important to underline the importance and increased interest in epidemiology studies on the epigenome of monozygotic twins, starting from the seminal paper of Fraga et al. in 2005 (Fraga et al. 2005).

Concerning seasonal allergic rhinitis (SAR), peripheral blood mononuclear cells from SAR patients and healthy controls were challenged with allergen. A different DNA methylation profile was observed in total $CD4^+$ T-cells for SAR patients compared to healthy controls allowing a classification based on DNA methylation. Interestingly, the DNA methylation profile of SAR patients varied during and outside pollen season (Nestor et al. 2014). In light of the importance of DNA methylation in inflammatory reactions, DNMTi were tested in vitro and in vivo for their ability to control inflammation. In asthma, upon allergen sensitization/challenge by ovalbumin treatment, a comparison of sensitized and non-sensitized mice showed an increase in *INFG* promoter methylation correlating with a decrease in INFγ level. This increase was reversed by 5azadC treatment (Brand et al. 2012).

In acute lung injury (ALI), an inflammatory lung disease mainly caused by sepsis, the combination of 5azadC and TSA reduced mortality level in lipopolysaccharide (LPS)-induced ALI mouse model with an 80 % survival rate (Thangavel et al. 2014). The same group reported that primary bone marrow-derived macrophages of LPS-treated mice expressed higher levels of pro-inflammatory chemokines and cytokines, which were decreased upon 5azadC/TSA treatment, resulting in cell survival increase. A dramatic synergy was observed from the combination compared to the single agents (Thangavel et al. 2015).

Rheumatoid arthritis is a chronic inflammatory autoimmune disease involving synovial inflammation and causing joint pains. In fibroblast-like synoviocytes (FLS) of arthritic rats, the methyl-CpG-binding protein 2 (MeCP2) was selectively overexpressed, whereas secreted frizzled-related protein 4 (SFRP4) was downregulated. A knockdown of MeCP2 resulted in an enhancement of the SFRP4 level. SFRP4 is known to activate the Wnt pathway, which is involved in FSL abnormal proliferation. Treatment of FLS with 5azadC resulted in an increased *SFRP4* expression and a decreased cell proliferation (Miao et al. 2013).

Finally, a strong increase in Th2 gene (*IL4* and *IL5*) promoter methylation with an opposite effect for Th1 genes (*INFG* and *IL10*) was recently observed in PMBC of children allergic to cow's milk compared to healthy ones (Berni Canani et al. 2015).

In summary, inflammation studies were carried out mainly with HDACi and few with DNMTi (Brook et al. 2015). The recent studies described above suggest the potential of DNMTi in the development of new anti-inflammatory treatment in asthma, acute lung injury, and rheumatoid arthritis.

3.4.4 Infection Diseases

Viral Infections

In 2014, 35 million people were affected by human immunodeficiency virus (HIV)-1 worldwide. Most infected people have a normal quality of life when treated with anti-retrovirus therapy, but an interruption in the treatment would reactivate latent viruses (Van Lint et al. 2013). Efforts have been made to improve the anti-HIV therapies. Proviruses, like HIV or human T-lymphotropic virus (HTLV), escape from the immune system and alter DNA methylation pattern (Saggioro et al. 1991) and more generally the epigenome (Kumar et al. 2015). In HIV-infected cells, increases of DNMT1 (Mikovits et al. 1998) and DNMT3A/DNMT3B levels (Chandel et al. 2015) and decrease in the expression of certain genes, such as *INFG*, *P16*, or vitamin D receptor gene, were shown (Fang et al. 2001; Chandel et al. 2015). However, the influence of DNA methylation is quite controversial, and a hypermethylation was reported in latent HIV reservoirs of patients who did not present viremia compared to viremic patients (Blazkova et al. 2009). A strategic option is to activate HIV in their reservoir to kill infected cells, while anti-retrovirals are used to block new infections. In this context, Fernandez et al. demonstrated that, in association with TNFα known to activate HIV replication, the use of 5azadC activated HIV twice more than the single use of TNFα in J-Lat cells. However, this seems to be cell line specific as in other cell lines, such as J1.1 and U1, and 5azadC

inhibited HIV activation (Fernandez and Zeichner 2010). This variability highlights the great need to optimize DNMTi use for HIV therapy.

HIV is also known to be associated with certain cancers such as non-Hodgkin's lymphomas (NHL). In HIV-positive NHL patients, a clear upregulation of DNMTs was observed. In HIV-positive aggressive B-cell lymphomas, the HIV-TAT protein was shown to be secreted by infected cells and to result in cell proliferation. In this context, Luzzi et al. used a model of TAT-transfected B-cell lymphoma cells and showed that this model upregulated DNMTs associated with aberrant hypermethylation. In this context, 5azadC was able to reexpress *P16* (Luzzi et al. 2014). Similarly, human papillomavirus (HPV)-positive lung cancers exhibited high levels of DNMT3B (Lin et al. 2005) and an upregulation of *E-cadherin*, a cell adhesion protein, associated with an increase of viral oncoprotein E6. Treatment with 5azadC was able to restore normal levels of *E-cadherin* expression in the presence of HPV (D'Costa et al. 2012). These findings could be exploited to improve the treatment of virus-induced cancers like HIV-positive NHL, papilloma virus-positive cancers, or Epstein–Barr virus-positive Burkitt's lymphoma (Paschos et al. 2009).

Hepatitis C virus (HCV) was also shown to lower E-cadherin levels and induce a hypermethylation of the *E-cadherin* promoter, which is known to induce morphological changes, alter cell–cell adhesion, and induce epithelial–mesenchymal transition, critical in tumorigenesis (Park and Jang 2014). DNA methylation was reported to be deregulated in HCV. DNMT activity was shown to be essential for HCV cell infection, and the use of 5aza or 5azadC induced a significant decrease in HCV infection (Chen et al. 2013). 5AzadC treatment of hematoma cells was shown to abolish HCV-induced *E-cadherin* downregulation and restore cell aggregation ability (Arora et al. 2008). Therefore, DNMTs can represent potential targets for HCV treatment.

Bacterial Infections

Epigenetic changes regulate bacteria-specific innate immune response, as described for inflammation (see Sect. 3.4.3). Alterations of DNMT expression levels were measured in mice provided with *E. coli* or *E. coli*-LPS-contaminated water compared to control mice. This highlighted the role of bacterial LPS in the alteration of the epigenetic response (Kovalchuk et al. 2013). In *E. coli*-infected human uroepithelial cells, DNMT activity was demonstrated to be more than tenfold enhanced compared to noninfected cells (Tolg et al. 2011). DNMT1 expression was also altered following *Porphyromonas gingivalis* or *Fusobacterium nucleatum* infection (Yin and Chung 2011). In parallel, pretreatment of gingival epithelial cells with 5aza significantly reduced the upregulation of cytokine genes *IL6* and *CXCL1*, observed in case of *P. gingivalis* or *F. nucleatum* exposure. However, no reduction in bacterial invasion of gingival epithelial cells was observed (Drury and Chung 2015). Hence, although DNA methylation is involved in bacterial infection, the potential influence of DNMTi is still to be studied.

The importance of the epigenetic modulation of the host in case of infective disease starts to be established, and this could open the door to the use of epidrugs as

anti-infectious agents. This would certainly be facilitated by the discovery of more potent and selective small molecule DNMTi.

3.4.5 Embryo Growth

The establishment and maintenance of DNA methylation patterns are crucial in the early development of the human embryo (Guo et al. 2014). Altered DNMT activity in the embryo can lead to early pregnancy loss (Yin et al. 2012). Early DNA methylation pattern establishment is also important in the in vitro fertilization (IVF) process, and development of preimplanted embryos was shown to be dependent on correct methylation (Dobbs et al. 2013). To better understand and improve the success rate in IVF process, studies were carried out on DNA methylation pattern in model animal embryos. *S*-Adenosyl-L-homocysteine (AdoHcy), described as a DNMTi based on product inhibition, was used to study in vitro mouse embryo development. Administered at a period of time preceding de novo methylation, AdoHcy results in a decrease of DNA methylation and improvement of development competency of in vitro cloned embryos compared to non-treated ones (Jafari et al. 2011). Treatment of buffalo skin fibroblast donor cells with the combination of 5azadC/TSA decreased global DNA methylation and apoptosis and improved the development of cloned embryos (Saini et al. 2014). RG108 in combination with an HDAC inhibitor showed synergetic effects on somatic cell nuclear transfer, a technique that is known to have limited efficiency in mammalian cloning and in which the DNA methylation profile was shown to be incompletely reprogrammed (Peat and Reik 2012). RG108/HDACi combination modulated the *XIST* methylation profile and reduced the expression of MDB3, which is overexpressed in cloned embryos (Xu et al. 2013). In another study, RG108 was evaluated as a single agent and at high doses (500 µM). It increased embryoblasts compared to non-treated controls, which is promising for the improvement in cloning efficiency (Li et al. 2011; Watanabe et al. 2013).

3.5 DNMTi Application in Microbial Production

To date, no direct therapeutic studies targeting DNA methylation in fungal infection were carried out. However, DNA methylation was found to be involved in fungal biosynthetic gene repression. When cultured in the presence of 50 µM of 5aza, Atlantic-forest-soil-derived *Penicillium citreonigrum* produced fungal exudates, formally known as guttates, which are secondary metabolites strongly enriched in several specific components, such as the azaphilone family members and new secondary metabolites from the atlantinone family (Wang et al. 2010a). Likewise, supplementation of the *Alternaria* sp. fungus culture medium by 5aza induced the production of toxic metabolites, such as alternariol, altenusin, alternariol-5-*O*-methyl ether, and 3′-hydroxyalternariol-5-*O*-methyl ether, which are known as plant disease inducers that are dormant in normal culture conditions (Sun et al. 2012).

Additionally, DNMTi and HDACi can be added to bacterial cultures to diversify the metabolite production. For example, 5aza was added to marine fungus *Leucostoma*

persoonii culture medium to optimize the production of cytosporones; two unknown cytosporones were obtained, one of which exhibiting some activity against *Plasmodium falciparum* (Beau et al. 2012). This modulation of the fungus biosynthesis pathways through DNA methylation tailoring could lead to the production and identification of novel molecules that can be therapeutically active compounds.

3.6 DNMTi in Plants

Variations in plant growth conditions, such as temperature, induce epimutations, which can lead to significant modifications in the phenotype, for example, the variation of petal number of the *Rosa hybrid*. This modification was related to DNA hypermethylation of the *RhAG* promoter at low temperature (Ma et al. 2015; Cortijo et al. 2014). In contrast to humans, cytosine methylation in plants does not occur only in a CpG context, and up to 30 % of the cytosines are methylated compared to 5–8 % in humans (Finnegan et al. 1998). Given the importance of DNA methylation in plants, DNMTi were tested in *Arabidopsis thaliana* (Zhang et al. 2006; Zilberman et al. 2007), tobacco (Vyskot et al. 1995), and rice. More precisely, in *Japonica* rice (*Oryza sativa L.*), 5aza and 5azadC (0.3 mM for 16 h to 3 days) were shown to decrease the DNA methylation level and to result in dwarfism, probably due to aberrant protein expression. The effect of 5azadC (70 % dwarfism effect) was more pronounced than 5aza (30 % dwarfism effect). Unlike in mammals, the DNA methylation profile is not erased during gametogenesis in plants (Kinoshita et al. 2004, 2007), which results in heritable DNA methylation changes in progenies, and 5aza-induced dwarfism in rice was shown to be maintained up to the third generation (Sano et al. 1990). Treatment of *Japonica* rice with 0.5 mM 5azadC caused a high lethality with less than 4 % germination. Among germinated seeds, a lower methylation at the *Xa21G* promoter was measured. *Xa21G* is a resistance gene to the pathogenic bacterium, *Xanthomonas oryzae pv. oryzae* (Ronald 1997), and, indeed, the new rice line obtained upon 5aza treatment was proven to possess a resistance trait toward bacterial infection (Akimoto et al. 2007). In another study, the combination of 5aza and a herbicide (2,4-dichlorophenoxyacetic acid) improved somatic embryogenesis in pineapple guava, *Acca sellowiana* (Fraga et al. 2012). In this case, 5aza was reported to counterbalance the hypermethylation induced by 2,4-dichlorophenoxyacetic acid.

Thus, DNMTi can program new features in plants, like the disease resistance trait described above or flowering (Kondo et al. 2007). However, as long as no plant-specific DNMTi is identified, attention must be taken that these treatments do not to affect the methylation profile of animals and human beings.

4 Limits and Perspectives

The above examples illustrate the major role of DNA methylation in the normal and abnormal functioning of cells and how its inhibition by DNMTi can change phenotypes, revealing the high potential of DNMTi, both for therapeutic strategies and for the bioengineering of organisms. It also highlights how, depending of the context,

the effects can be inversed. This is related to the plasticity of DNA methylation and its role in controlling gene expression. As illustrated by the above examples, DNMTi are just starting to be applied to other diseases beyond cancer, and these strategies seem very promising. However, cancer treatment is still today the most studied medical application of DNMTi.

In the past couple of years, great hopes came from the use of 5aza and 5azadC to reprogram cancer cells toward increased chemosensitivity (Clozel et al. 2013) or sensitize them toward immunotherapy (Azad et al. 2013). This means that it might be possible to overpass chemoresistance, which is a big burden in cancer management, and to open immunotherapies to nonresponding patients. An increasing number of clinical trials are starting to explore this epigenetic reprogramming in particular in solid tumors. Time will show if these studies can validate the hypothesis.

Noteworthy, the only potent DNMTi currently available are 5aza and 5azadC, which are suicide substrates that are incorporated into DNA and form an irreversible covalent complex with the enzyme. This triggers a subsequent DNMT degradation by the proteasome, inducing a potent inhibition of DNA methylation (Erdmann et al. 2015). It is not clear whether a non-covalent inhibitor can induce this level of demethylation. Can we thus have the same potency with non-nucleoside inhibitors? Certainly such inhibitors have to present a very high affinity for the DNMT. Maybe covalent non-nucleoside inhibitors could be an alternative, but they would need to be very specific for the DNMTs and today this specificity is lacking. Thus, it is urgent to find novel compounds that are very potent inhibitors of DNMTs and bind strongly to the enzyme, in order to induce a strong inhibition in cells and eventually a reduction in DNA methylation.

Another feature of 5aza, 5azadC, and their analogs is that they are incorporated in the DNA instead of all deoxycytidine and thus they are not specific for CpGs or certain genomic regions. This can hinder their use by inducing non-desired secondary effects. The use of low doses of the drugs has diminished these effects; however, repeated cycles are necessary for the epigenetic reprogramming, increasing the probability of their appearance. Non-nucleoside inhibitors have the advantage of not needing to be incorporated into DNA and thus diminish the side effects. In conclusion, it is important to pursue the search for new non-nucleoside inhibitors of DNA methylation that are potent, specific for the CpGs and, in particular, of promoters silencing key genes in pathologies. We have recently discussed several strategies that can be explored for the next generation of inhibitors, as allosteric inhibitors, protein–protein ligands, and dual inhibitors (Erdmann et al. 2016).

References

Ahuja N, Easwaran H, Baylin SB. Harnessing the potential of epigenetic therapy to target solid tumors. J Clin Invest. 2014;124(1):56–63. doi:10.1172/jci69736.

Ahuja N, Sharma AR, Baylin SB. Epigenetic therapeutics: a new weapon in the war against cancer. Annu Rev Med. 2016;67(1):73–89. doi:10.1146/annurev-med-111314-035900.

Aizawa SY. Yutaka. Valproate administration to mice increases hippocampal p21 expression by altering genomic DNA methylation. Neuroreport. 2015;26(15):915–20. doi:10.1097/WNR.0000000000000448.

Akimoto K, Katakami H, Kim H-J, Ogawa E, Sano CM, Wada Y, et al. Epigenetic inheritance in rice plants. Ann Bot. 2007;100(2):205–17. doi:10.1093/aob/mcm110.
Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat Genet. 1999;23(2):185–8.
Arora P, Kim E-O, Jung JK, Jang KL. Hepatitis C virus core protein downregulates E-cadherin expression via activation of DNA methyltransferase 1 and 3b. Cancer Lett. 2008;261(2):244–52. http://dx.doi.org/10.1016/j.canlet.2007.11.033.
Asgatay S, Champion C, Marloie G, Drujon T, Senamaud-Beaufort C, Ceccaldi A, et al. Synthesis and evaluation of analogues of *N*-phthaloyl-L-tryptophan (RG108) as inhibitors of DNA methyltransferase 1. J Med Chem. 2014;57(2):421–34. doi:10.1021/jm401419p.
Aydin B, Hocaoglu N, Gidener S. Could decitabine treatment impair memory functions in humans? Med Hypotheses. 2012;79(5):639–41. http://dx.doi.org/10.1016/j.mehy.2012.07.042.
Azad N, Zahnow CA, Rudin CM, Baylin SB. The future of epigenetic therapy in solid tumours—lessons from the past. Nat Rev Clin Oncol. 2013;10(5):256–66. doi:10.1038/nrclinonc.2013.42.
Barbier E, Tapocik JD, Juergens N, Pitcairn C, Borich A, Schank JR, et al. DNA methylation in the medial prefrontal cortex regulates alcohol-induced behavior and plasticity. J Neurosci. 2015;35(15):6153–64. doi:10.1523/jneurosci.4571-14.2015.
Baud MG, Leiser T, Haus P, Samlal S, Wong AC, Wood RJ, et al. Defining the mechanism of action and enzymatic selectivity of psammaplin A against its epigenetic targets. J Med Chem. 2012;55(4):1731–50. doi:10.1021/jm2016182.
Beau J, Mahid N, Burda WN, Harrington L, Shaw LN, Mutka T, et al. Epigenetic tailoring for the production of anti-infective cytosporones from the *Marine Fungus Leucostoma persoonii*. Mar Drugs. 2012;10(4):762–74. doi:10.3390/md10040762.
Benton M, Johnstone A, Eccles D, Harmon B, Hayes M, Lea R, et al. An analysis of DNA methylation in human adipose tissue reveals differential modification of obesity genes before and after gastric bypass and weight loss. Genome Biol. 2015;16(1):8.
Berni Canani R, Paparo L, Nocerino R, Cosenza L, Pezzella V, Di Costanzo M, et al. Differences in DNA methylation profile of Th1 and Th2 cytokine genes are associated with tolerance acquisition in children with IgE-mediated cow's milk allergy. Clin Epigenet. 2015;7(1):38. doi:10.1186/s13148-015-0070-8.
Blazkova J, Trejbalova K, Gondois-Rey F, Halfon P, Philibert P, Guiguen A, et al. CpG methylation controls reactivation of HIV from latency. PLoS Pathog. 2009;5(8):e1000554. doi:10.1371/journal.ppat.1000554.
Boku S, Toda H, Nakagawa S, Kato A, Inoue T, Koyama T, et al. Neonatal maternal separation alters the capacity of adult neural precursor cells to differentiate into neurons via methylation of retinoic acid receptor gene promoter. Biol Psychiatry. 2015;77(4):335–44. http://dx.doi.org/10.1016/j.biopsych.2014.07.008.
Bönsch D, Lenz B, Fiszer R, Frieling H, Kornhuber J, Bleich S. Lowered DNA methyltransferase (DNMT-3b) mRNA expression is associated with genomic DNA hypermethylation in patients with chronic alcoholism. J Neural Transm. 2006;113(9):1299–304. doi:10.1007/s00702-005-0413-2.
Booij L, Wang D, Lévesque ML, Tremblay RE, Szyf M. Looking beyond the DNA sequence: the relevance of DNA methylation processes for the stress–diathesis model of depression. Philos Trans R Soc Lond B Biol Sci. 2013;368(1615). doi:10.1098/rstb.2012.0251.
Boothman DA, Briggle TV, Greer S. Exploitation of elevated pyrimidine deaminating enzymes for selective chemotherapy. Pharmacol Ther. 1989;42(1):65–88.
Bouchard L, Rabasa-Lhoret R, Faraj M, Lavoie M-È, Mill J, Pérusse L, et al. Differential epigenomic and transcriptomic responses in subcutaneous adipose tissue between low and high responders to caloric restriction. Am J Clin Nutr. 2010;91(2):309–20. doi:10.3945/ajcn.2009.28085.
Brand S, Kesper DA, Teich R, Kilic-Niebergall E, Pinkenburg O, Bothur E, et al. DNA methylation of TH1/TH2 cytokine genes affects sensitization and progress of experimental asthma. J Allergy Clin Immunol. 2012;129(6):1602–10. doi:10.1016/j.jaci.2011.12.963.
Brøns C, Jacobsen S, Nilsson E, Rönn T, Jensen CB, Storgaard H, et al. Deoxyribonucleic acid methylation and gene expression of PPARGC1A in human muscle is influenced by high-fat

overfeeding in a birth-weight-dependent manner. J Clin Endocrinol Metab. 2010;95(6):3048–56. doi:10.1210/jc.2009-2413.

Brook PO, Perry MM, Adcock IM, Durham AL. Epigenome-modifying tools in asthma. Epigenomics. 2015:[Epub ahead of print]. doi:10.2217/epi.15.53.

Brueckner B, Garcia Boy R, Siedlecki P, Musch T, Kliem HC, Zielenkiewicz P, et al. Epigenetic reactivation of tumor suppressor genes by a novel small-molecule inhibitor of human DNA methyltransferases. Cancer Res. 2005;65(14):6305–11. doi:10.1158/0008-5472.can-04-2957.

Brueckner B, Rius M, Markelova MR, Fichtner I, Hals PA, Sandvold ML, et al. Delivery of 5-azacytidine to human cancer cells by elaidic acid esterification increases therapeutic drug efficacy. Mol Cancer Ther. 2010;9(5):1256–64. doi:10.1158/1535-7163.MCT-09-1202.

Byun HM, Choi SH, Laird PW, Trinh B, Siddiqui MA, Marquez VE, et al. 2′-Deoxy-N4-[2-(4-nitrophenyl)ethoxycarbonyl]-5-azacytidine: a novel inhibitor of DNA methyltransferase that requires activation by human carboxylesterase 1. Cancer Lett. 2008;266(2):238–48. doi:10.1016/j.canlet.2008.02.069.

Cameron EE, Bachman KE, Myohanen S, Herman JG, Baylin SB. Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer. Nat Genet. 1999;21(1):103–7.

Candelaria M, Herrera A, Labardini J, González-Fierro A, Trejo-Becerril C, Taja-Chayeb L, et al. Hydralazine and magnesium valproate as epigenetic treatment for myelodysplastic syndrome. Preliminary results of a phase-II trial. Ann Hematol. 2011;90(4):379–87. doi:10.1007/s00277-010-1090-2.

Candelaria M, de la Cruz-Hernandez E, Taja-Chayeb L, Perez-Cardenas E, Trejo-Becerril C, Gonzalez-Fierro A, et al. DNA methylation-independent reversion of gemcitabine resistance by hydralazine in cervical cancer cells. PLoS One. 2012;7(3):e29181. doi:10.1371/journal.pone.0029181.

Carnicella S, Kharazia V, Jeanblanc J, Janak PH, Ron D. GDNF is a fast-acting potent inhibitor of alcohol consumption and relapse. Proc Natl Acad Sci U S A. 2008;105(23):8114–9. doi:10.1073/pnas.0711755105.

Castellano S, Kuck D, Sala M, Novellino E, Lyko F, Sbardella G. Constrained analogues of procaine as novel small molecule inhibitors of DNA methyltransferase-1. J Med Chem. 2008;51(7):2321–5. doi:10.1021/jm7015705.

Castellano S, Kuck D, Viviano M, Yoo J, Lopez-Vallejo F, Conti P, et al. Synthesis and biochemical evaluation of Δ2-isoxazoline derivatives as DNA methyltransferase 1 inhibitors. J Med Chem. 2011;54(21):7663–77.

Ceccaldi A, Rajavelu A, Champion C, Rampon C, Jurkowska R, Jankevicius G, et al. C5-DNA methyltransferase inhibitors: from screening to effects on zebrafish embryo development. Chembiochem. 2011;12(9):1337–45.

Ceccaldi A, Rajavelu A, Ragozin S, Sénamaud-Beaufort C, Bashtrykov P, Testa N, et al. Identification of novel inhibitors of DNA methylation by screening of a chemical library. ACS Chem Biol. 2013;8(3):543–8.

Champion C, Guianvarc'h D, Sénamaud-Beaufort C, Jurkowska RZ, Jeltsch A, Ponger L, et al. Mechanistic insights on the inhibition of C5 DNA methyltransferases by zebularine. PLoS One. 2010;5(8):e12388. doi:10.1371/journal.pone.0012388.

Chan KK, Giannini DD, Staroscik JA, Sadee W. 5-azacytidine hydrolysis kinetics measured by high-pressure liquid chromatography and 13C-NMR spectroscopy. J Pharm Sci. 1979;68(7):807–12. doi:10.1002/jps.2600680705.

Chandel N, Ayasolla KS, Lan X, Sultana-Syed M, Chawla A, Lederman R, et al. Epigenetic modulation of human podocyte vitamin D receptor in HIV Milieu. J Mol Biol. 2015;427(20):3201–15. http://dx.doi.org/10.1016/j.jmb.2015.07.011.

Chaturvedi P, Tyagi SC. Epigenetic mechanisms underlying cardiac degeneration and regeneration. Int J Cardiol. 2014;173(1):1–11. http://dx.doi.org/10.1016/j.ijcard.2014.02.008.

Chen Y, Sharma RP, Costa RH, Costa E, Grayson DR. On the epigenetic regulation of the human reelin promoter. Nucleic Acids Res. 2002;30(13):2930–9. doi:10.1093/nar/gkf401.

Chen C, Pan D, Deng AM, Huang F, Sun B-L, Yang R-G. DNA methyltransferases 1 and 3B are required for hepatitis C virus infection in cell culture. Virology. 2013;441(1):57–65. http://dx.doi.org/10.1016/j.virol.2013.03.005.

Cheng JC, Matsen CB, Gonzales FA, Ye W, Greer S, Marquez VE, et al. Inhibition of DNA methylation and reactivation of silenced genes by zebularine. J Natl Cancer Inst. 2003;95(5): 399–409.

Chestnut BA, Chang Q, Price A, Lesuisse C, Wong M, Martin LJ. Epigenetic regulation of motor neuron cell death through DNA methylation. J Neurosci. 2011;31(46):16619–36. doi:10.1523/jneurosci.1639-11.2011.

Cho G-W, Noh M-Y, Kim HY, Koh S-H, Kim K-S, Kim SH. Bone marrow-derived stromal cells from amyotrophic lateral sclerosis patients have diminished stem cell capacity. Stem Cells Dev. 2010;19(7):1035–42. doi:10.1089/scd.2009.0453.

Chuang JC, Yoo CB, Kwan JM, Li TWH, Liang G, Yang AS, et al. Comparison of biological effects of non-nucleoside DNA methylation inhibitors versus 5-aza-2′-deoxycytidine. Mol Cancer Ther. 2005;4(10):1515–20. doi:10.1158/1535-7163.mct-05-0172.

Chuang JC, Warner SL, Vollmer D, Vankayalapati H, Redkar S, Bearss DJ, et al. S110, a 5-Aza-2′-deoxycytidine–containing dinucleotide, is an effective DNA methylation inhibitor *in vivo* and can reduce tumor growth. Mol Cancer Ther. 2010;9(5):1443–50. doi:10.1158/1535-7163.mct-09-1048.

Cihak A. Biological effects of 5-azacytidine in eukaryotes. Oncology. 1974;30(5):405–22.

Cleveland DW, Rothstein JD. From Charcot to Lou Gehrig: deciphering selective motor neuron death in ALS. Nat Rev Neurosci. 2001;2(11):806–19.

Clozel T, Yang S, Elstrom RL, Tam W, Martin P, Kormaksson M, et al. Mechanism-based epigenetic chemosensitization therapy of diffuse large B-cell lymphoma. Cancer Discov. 2013;3(9):1002–19. doi:10.1158/2159-8290.cd-13-0117.

Constantinides PG, Jones PA, Gevers W. Functional striated muscle cells from non-myoblast precursors following 5-azacytidine treatment. Nature. 1977;267(5609):364–6.

Constantinides PG, Taylor SM, Jones PA. Phenotypic conversion of cultured mouse embryo cells by aza pyrimidine nucleosides. Dev Biol. 1978;66(1):57–71. http://dx.doi.org/10.1016/0012-1606(78)90273-7.

Coppieters N, Dragunow M. Epigenetics in alzheimers disease: a focus on DNA modifications. Curr Pharm Des. 2011;17(31):3398–412. doi:10.2174/138161211798072544#sthash.ZNP9Kotu.dpuf.

Coppieters N, Dieriks BV, Lill C, Faull RLM, Curtis MA, Dragunow M. Global changes in DNA methylation and hydroxymethylation in Alzheimer's disease human brain. Neurobiol Aging. 2014;35(6):1334–44. http://dx.doi.org/10.1016/j.neurobiolaging.2013.11.031.

Cornacchia E, Golbus J, Maybaum J, Strahler J, Hanash S, Richardson B. Hydralazine and procainamide inhibit T cell DNA methylation and induce autoreactivity. J Immunol. 1988; 140(7):2197–200.

Cortijo S, Wardenaar R, Colomé-Tatché M, Gilly A, Etcheverry M, Labadie K, et al. Mapping the epigenetic basis of complex traits. Science. 2014;343(6175):1145–8. doi:10.1126/science.1248127.

Crescenti A, Solà R, Valls RM, Caimari A, del Bas JM, Anguera A, et al. Cocoa consumption alters the global DNA methylation of peripheral leukocytes in humans with cardiovascular disease risk factors: a randomized controlled trial. PLoS One. 2013;8(6):e65744. doi:10.1371/journal.pone.0065744.

D'Costa ZJ, Jolly C, Androphy EJ, Mercer A, Matthews CM, Hibma MH. Transcriptional repression of E-cadherin by human papillomavirus type 16 E6. PLoS One. 2012;7(11):e48954. doi:10.1371/journal.pone.0048954.

Datta J, Ghoshal K, Denny WA, Gamage SA, Brooke DG, Phiasivongsa P, et al. A new class of quinoline-based DNA hypomethylating agents reactivates tumor suppressor genes by blocking DNA methyltransferase 1 activity and inducing its degradation. Cancer Res. 2009;69(10):4277–85. doi:10.1158/0008-5472.can-08-3669.

Day JJ, Sweatt JD. DNA methylation and memory formation. Nat Neurosci. 2010;13(11): 1319–23.

De Jager PL, Srivastava G, Lunnon K, Burgess J, Schalkwyk LC, Yu L, et al. Alzheimer's disease: early alterations in brain DNA methylation at *ANK1*, *BIN1*, *RHBDF2* and other loci. Nat Neurosci. 2014;17(9):1156–63. doi:10.1038/nn.3786. http://www.nature.com/neuro/journal/v17/n9/abs/nn.3786.html#supplementary-information.

de la Cruz-Hernandez E, Perez-Cardenas E, Taja-Chayeb L, Chavez-Blanco A, Trejo-Becerril C, Trujillo J, et al. DNA demethylating activity of hydralazine in cancer cell lines. Life Sci Med Res. 2011;2011:LSMR-30.

Desplats P, Spencer B, Coffee E, Patel P, Michael S, Patrick C, et al. α-synuclein sequesters Dnmt1 from the nucleus: a novel mechanism for epigenetic alterations in Lewy body diseases. J Biol Chem. 2011;286(11):9031–7. doi:10.1074/jbc.C110.212589.

Ding L, Qiu L, Zhang J, Guo B. Camptothecin-induced cell proliferation inhibition and apoptosis enhanced by DNA methyltransferase inhibitor, 5-Aza-2′-deoxycytidine. Biol Pharm Bull. 2009;32(6):1105–8. doi:10.1248/bpb.32.1105.

Dobbs KB, Rodriguez M, Sudano MJ, Ortega MS, Hansen PJ. Dynamics of DNA methylation during early development of the preimplantation bovine embryo. PLoS One. 2013;8(6):e66230. doi:10.1371/journal.pone.0066230.

Drury JL, Chung WO. DNA methylation differentially regulates cytokine secretion in gingival epithelia in response to bacterial challenges. Pathog Dis. 2015;73(2):1–6. doi:10.1093/femspd/ftu005.

Duenas-Gonzalez A, Candelaria M, Perez-Plascencia C, Perez-Cardenas E, de la Cruz-Hernandez E, Herrera LA. Valproic acid as epigenetic cancer drug: preclinical, clinical and transcriptional effects on solid tumors. Cancer Treat Rev. 2008;34(3):206–22. http://dx.doi.org/10.1016/j.ctrv.2007.11.003.

Dueñas-Gonzalez A, Coronel J, Cetina L, González-Fierro A, Chavez-Blanco A, Taja-Chayeb L. Hydralazine–valproate: a repositioned drug combination for the epigenetic therapy of cancer. Expert Opin Drug Metab Toxicol. 2014;10(10):1433–44. doi:10.1517/17425255.2014.947263.

Dunn J, Qiu H, Kim S, Jjingo D, Hoffman R, Kim CW, et al. Flow-dependent epigenetic DNA methylation regulates endothelial gene expression and atherosclerosis. J Clin Invest. 2014;124(7):3187–99. doi:10.1172/jci74792.

Erdmann A, Halby L, Fahy J, Arimondo PB. Targeting DNA methylation with small molecules: what's next? J Med Chem. 2015;58(6):2569–83. doi:10.1021/jm500843d.

Erdmann A, Guianvarc'h D, Arimondo PB. Structure-guided optimization of DNA methyltransferase inhibitors. In: Medina-Franco J, editor. Epi-informatics, discovery and development of small molecule epigenetic drugs and probes. 1st ed. Elsevier, Future Med Chem, Future Science Group. 2016;8(4):373–80.

Ernst C, Deleva V, Deng X, Sequeira A, Pomarenski A, Klempan T, et al. Alternative splicing, methylation state, and expression profile of tropomyosin-related kinase B in the frontal cortex of suicide completers. Arch Gen Psychiatry. 2009;66(1):22–32. doi:10.1001/archpsyc.66.1.22.

Esteller M. Epigenetics in cancer. N Engl J Med. 2008;358(11):1148–59. doi:10.1056/NEJMra072067.

Evans JS, Hanka LJ. The *in vivo* activity of combinations of 5-azacytidine and cytidine on leukemia L-1210. Experientia. 1968;24(9):922–3.

Fagan RL, Cryderman DE, Kopelovich L, Wallrath LL, Brenner C. Laccaic Acid A is a direct, DNA-competitive inhibitor of DNA methyltransferase 1. J Biol Chem. 2013;288(33):23858–67.

Fahy J, Jeltsch A, Arimondo PB. DNA methyltransferase inhibitors in cancer: a chemical and therapeutic patent overview and selected clinical studies. Expert Opin Ther Pat. 2012;22(12):1427–42. doi:10.1517/13543776.2012.729579.

Fandy TE, Herman JG, Kerns P, Jiemjit A, Sugar EA, Choi S-H, et al. Early epigenetic changes and DNA damage do not predict clinical response in an overlapping schedule of 5-azacytidine and entinostat in patients with myeloid malignancies. Blood. 2009;114(13):2764–73. doi:10.1182/blood-2009-02-203547.

Fang J-Y, Mikovits JA, Bagni R, Petrow-Sadowski CL, Ruscetti FW. Infection of lymphoid cells by integration-defective human immunodeficiency virus type 1 increases *De Novo* methylation. J Virol. 2001;75(20):9753–61. doi:10.1128/jvi.75.20.9753-9761.2001.

Fang MZ, Wang Y, Ai N, Hou Z, Sun Y, Lu H, et al. Tea polyphenol (−)-epigallocatechin-3-gallate inhibits DNA methyltransferase and reactivates methylation-silenced genes in cancer cell lines. Cancer Res. 2003;63(22):7563–70.

Fang MZ, Chen D, Sun Y, Jin Z, Christman JK, Yang CS. Reversal of hypermethylation and reactivation of p16INK4a, RARβ, and MGMT genes by genistein and other isoflavones from Soy. Clin Cancer Res. 2005;11(19):7033–41. doi:10.1158/1078-0432.ccr-05-0406.

Feng J, Zhou Y, Campbell SL, Le T, Li E, Sweatt JD, et al. Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. Nat Neurosci. 2010;13(4):423–30. http://www.nature.com/neuro/journal/v13/n4/suppinfo/nn.2514_S1.html.

Fernandez G, Zeichner S. Cell line-dependent variability in HIV activation employing DNMT inhibitors. Virol J. 2010;7(1):266. doi:10.1186/1743-422X-7-266.

Finnegan EJ, Genger RK, Peacock WJ, Dennis ES. DNA methylation in plants. Annu Rev Plant Physiol Plant Mol Biol. 1998;49(1):223–47. doi:10.1146/annurev.arplant.49.1.223.

Flotho C, Claus R, Batz C, Schneider M, Sandrock I, Ihde S, et al. The DNA methyltransferase inhibitors azacitidine, decitabine and zebularine exert differential effects on cancer gene expression in acute myeloid leukemia cells. Leukemia. 2009;23(6):1019–28. http://www.nature.com/leu/journal/v23/n6/suppinfo/leu2008397s1.html.

Fraga MF, Ballestar E, Paz MF, Ropero S, Setien F, Ballestar ML, et al. Epigenetic differences arise during the lifetime of monozygotic twins. Proc Natl Acad Sci U S A. 2005;102(30):10604–9. doi:10.1073/pnas.0500398102.

Fraga HPF, Vieira LN, Caprestano CA, Steinmacher DA, Micke GA, Spudeit DA, et al. 5-Azacytidine combined with 2,4-D improves somatic embryogenesis of *Acca sellowiana* (O. Berg) Burret by means of changes in global DNA methylation levels. Plant Cell Rep. 2012;31(12):2165–76. doi:10.1007/s00299-012-1327-8.

Friso S, Pizzolo F, Choi S-W, Guarini P, Castagna A, Ravagnani V, et al. Epigenetic control of 11 beta-hydroxysteroid dehydrogenase 2 gene promoter is related to human hypertension. Atherosclerosis. 2008;199(2):323–7. http://dx.doi.org/10.1016/j.atherosclerosis.2007.11.029.

Gamage SA, Brooke DG, Redkar S, Datta J, Jacob ST, Denny WA. Structure-activity relationships for 4-anilinoquinoline derivatives as inhibitors of the DNA methyltransferase enzyme DNMT1. Bioorg Med Chem. 2013;21(11):3147–53. doi:10.1016/j.bmc.2013.03.033.

Gervin K, Vigeland MD, Mattingsdal M, Hammerø M, Nygård H, Olsen AO, et al. DNA methylation and gene expression changes in monozygotic twins discordant for psoriasis: identification of epigenetically dysregulated genes. PLoS Genet. 2012;8(1):e1002454. doi:10.1371/journal.pgen.1002454.

Graça I, Sousa EJ, Baptista T, Almeida M, Ramalho-Carvalho J, Palmeira C, et al. Anti-tumoral effect of the non-nucleoside DNMT inhibitor RG108 in human prostate cancer cells. Curr Pharm Des. 2014a;20(11):1803–11.

Graça I, Sousa EJ, Costa-Pinheiro P, Vieira FQ, Torres-Ferreira J, Martins MG, et al. Anti-neoplastic properties of hydralazine in prostate cancer. Oncotarget. 2014b;5(15):5950–64.

Grayson DR, Jia X, Chen Y, Sharma RP, Mitchell CP, Guidotti A, et al. Reelin promoter hypermethylation in schizophrenia. Proc Natl Acad Sci U S A. 2005;102(26):9341–6. doi:10.1073/pnas.0503736102.

Gros C, Fleury L, Nahoum V, Faux C, Valente S, Labella D, et al. New insights on the mechanism of quinoline-based DNA methyltransferase inhibitors. J Biol Chem. 2015;290(10):6293–302. doi:10.1074/jbc.M114.594671.

Guay S-P, Legare C, Houde A-A, Mathieu P, Bosse Y, Bouchard L. Acetylsalicylic acid, aging and coronary artery disease are associated with ABCA1 DNA methylation in men. Clin Epigenet. 2014;6(1):14.

Guidotti A, Auta J, Davis JM, DiGiorgi GV, Dwivedi Y, Grayson DR, et al. Decrease in Reelin and Glutamic Acid Decarboxylase67 (GAD67) expression in schizophrenia and bipolar disorder: a postmortem brain study FREE. Arch Gen Psychiatry. 2000;57(11):1061–9. doi:10.1001/archpsyc.57.11.1061.

Guidotti A, Dong E, Kundakovic M, Satta R, Grayson DR, Costa E. Characterization of the action of antipsychotic subtypes on valproate-induced chromatin remodeling. Trends Pharmacol Sci. 2009;30(2):55–60. http://dx.doi.org/10.1016/j.tips.2008.10.010.

Guidotti A, Auta J, Chen Y, Davis JM, Dong E, Gavin DP, et al. Epigenetic GABAergic targets in schizophrenia and bipolar disorder. Neuropharmacology. 2011;60(7–8):1007–16. http://dx.doi.org/10.1016/j.neuropharm.2010.10.021.

Guo H, Zhu P, Yan L, Li R, Hu B, Lian Y, et al. The DNA methylation landscape of human early embryos. Nature. 2014;511(7511):606–10. doi:10.1038/nature13544, http://www.nature.com/nature/journal/v511/n7511/abs/nature13544.html#supplementary-information.

Guy J, Gan J, Selfridge J, Cobb S, Bird A. Reversal of neurological defects in a mouse model of Rett Syndrome. Science. 2007;315(5815):1143–7. doi:10.1126/science.1138389.

Halby L, Champion C, Sénamaud-Beaufort C, Ajjan S, Drujon T, Rajavelu A, et al. Rapid synthesis of New DNMT inhibitors derivatives of procainamide. Chembiochem. 2012;13(1):157–65.

Heinrich C, Nitta N, Flubacher A, Müller M, Fahrner A, Kirsch M, et al. Reelin deficiency and displacement of mature neurons, but not neurogenesis, underlie the formation of granule cell dispersion in the epileptic hippocampus. J Neurosci. 2006;26(17):4701–13. doi:10.1523/jneurosci.5516-05.2006.

Horrillo A, Pezzolla D, Fraga MF, Aguilera Y, Salguero-Aranda C, Tejedo JR, et al. Zebularine regulates early stages of mESC differentiation: effect on cardiac commitment. Cell Death Dis. 2013;4:e570. doi:10.1038/cddis.2013.88.

Hossain M, Healey M, Lee C, Poh W, Yerram S, Patel K, et al. DNA-intercalators causing rapid re-expression of methylated and silenced genes in cancer cells. Oncotarget. 2013;4(2):298–309.

Issa J-PJ, Kantarjian HM. Targeting DNA methylation. Clin Cancer Res. 2009;15(12):3938–46. doi:10.1158/1078-0432.ccr-08-2783.

Jafari S, Hosseini MS, Hajian M, Forouzanfar M, Jafarpour F, Abedi P, et al. Improved *in vitro* development of cloned bovine embryos using *S*-adenosylhomocysteine, a non-toxic epigenetic modifying reagent. Mol Reprod Dev. 2011;78(8):576–84. doi:10.1002/mrd.21344.

Jones PA, Taylor SM. Cellular differentiation, cytidine analogs and DNA methylation. Cell. 1980;20(1):85–93. http://dx.doi.org/10.1016/0092-8674(80)90237-8.

Jowaed A, Schmitt I, Kaut O, Wüllner U. Methylation regulates α-synuclein expression and is decreased in Parkinson's disease patients' brains. J Neurosci. 2010;30(18):6355–9. doi:10.1523/jneurosci.6119-09.2010.

Juergens RA, Wrangle J, Vendetti FP, Murphy SC, Zhao M, Coleman B, et al. Combination epigenetic therapy has efficacy in patients with refractory advanced non-small cell lung cancer. Cancer Discov. 2011;1(7):OF1–10. doi:10.1158/2159-8290.cd-11-0214.

Kadowaki T, Yamauchi T, Kubota N, Hara K, Ueki K, Tobe K. Adiponectin and adiponectin receptors in insulin resistance, diabetes, and the metabolic syndrome. J Clin Invest. 2006;116(7):1784–92. doi:10.1172/jci29126.

Kang H-J, Kim J-M, Lee J-Y, Kim S-Y, Bae K-Y, Kim S-W, et al. *BDNF* promoter methylation and suicidal behavior in depressive patients. J Affect Disord. 2013;151(2):679–85. http://dx.doi.org/10.1016/j.jad.2013.08.001.

Kao Y-H, Chen Y-C, Chung C-C, Lien G-S, Chen S-A, Kuo C-C, et al. Heart failure and angiotensin II modulate atrial *Pitx2c* promotor methylation. Clin Exp Pharmacol Physiol. 2013;40(6):379–84. doi:10.1111/1440-1681.12089.

Karahoca M, Momparler RL. Pharmacokinetic and pharmacodynamic analysis of 5-aza-2′-deoxycytidine (decitabine) in the design of its dose-schedule for cancer therapy. Clin Epigenet. 2013;5(1):3. doi:10.1186/1868-7083-5-3.

Kilgore JA, Du X, Melito L, Wei S, Wang C, Chin HG, et al. Identification of DNMT1 selective antagonists using a novel scintillation proximity assay. J Biol Chem. 2013;288(27):19673–84. doi:10.1074/jbc.M112.443895.

Kim J, Kim JY, Song KS, Lee YH, Seo JS, Jelinek J et al. Epigenetic changes in estrogen receptor β gene in atherosclerotic cardiovascular tissues and in-vitro vascular senescence. Biochim Biophys Acta Mol Basis Dis. 2007;1772(1):72–80. http://dx.doi.org/10.1016/j.bbadis.2006.10.004.

Kim M, Long TI, Arakawa K, Wang R, Yu MC, Laird PW. DNA methylation as a biomarker for cardiovascular disease risk. PLoS One. 2010;5(3):e9692. doi:10.1371/journal.pone.0009692.

Kim HJ, Kim JH, Chie EK, Da Young P, Kim IA, Kim IH. DNMT (DNA methyltransferase) inhibitors radiosensitize human cancer cells by suppressing DNA repair activity. Radiat Oncol. 2012;7(1):39.

Kim AY, Park YJ, Pan X, Shin KC, Kwak S-H, Bassas AF, et al. Obesity-induced DNA hypermethylation of the adiponectin gene mediates insulin resistance. Nat Commun. 2015;6:7585. doi:10.1038/ncomms8585.

Kinoshita T, Miura A, Choi Y, Kinoshita Y, Cao X, Jacobsen SE, et al. One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. Science. 2004;303(5657):521–3. doi:10.1126/science.1089835.

Kinoshita Y, Saze H, Kinoshita T, Miura A, Soppe WJJ, Koornneef M, et al. Control of FWA gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats. Plant J. 2007;49(1):38–45. doi:10.1111/j.1365-313X.2006.02936.x.

Kobow K, Jeske I, Hildebrandt M, Hauke J, Hahnen E, Buslei R, et al. Increased reelin promoter methylation is associated with granule cell dispersion in human temporal lobe epilepsy. J Neuropathol Exp Neurol. 2009;68(4):356–64. doi:10.1097/NEN.0b013e31819ba737.

Kobow K, Kaspi A, Harikrishnan KN, Kiese K, Ziemann M, Khurana I, et al. Deep sequencing reveals increased DNA methylation in chronic rat epilepsy. Acta Neuropathol. 2013;126(5):741–56. doi:10.1007/s00401-013-1168-8.

Kohli A, Garcia M, Miller R, Maher C, Humblet O, Hammond S, et al. Secondhand smoke in combination with ambient air pollution exposure is associated with increased CpG methylation and decreased expression of IFN-γ in T effector cells and Foxp3 in T regulatory cells in children. Clin Epigenet. 2012;4(1):17.

Kondo H, Miura T, Wada KC, Takeno K. Induction of flowering by 5-azacytidine in some plant species: relationship between the stability of photoperiodically induced flowering and flower-inducing effect of DNA demethylation. Physiol Plant. 2007;131(3):462–9. doi:10.1111/j.1399-3054.2007.00965.x.

Kovalchuk I, Walz P, Thomas J, Kovalchuk O. The increased expression of proteins involved in proliferation, DNA repair and DNA methylation in spleen of mice exposed to E. coli O157:H7 lipopolysaccharide. Environ Mol Mutagen. 2013;54(6):421–8. doi:10.1002/em.21787.

Kuck D, Caulfield T, Lyko F, Medina-Franco JL. Nanaomycin A selectively inhibits DNMT3B and reactivates silenced tumor suppressor genes in human cancer cells. Mol Cancer Ther. 2010;9(11):3015–23. doi:10.1158/1535-7163.mct-10-0609.

Kumar A, Darcis G, Van Lint C, Herbein G. Epigenetic control of HIV-1 post integration latency: implications for therapy. Clin Epigenet. 2015;7(1):103. doi:10.1186/s13148-015-0137-6.

Kundakovic M, Chen Y, Costa E, Grayson DR. DNA methyltransferase inhibitors coordinately induce expression of the human reelin and glutamic acid decarboxylase 67 genes. Mol Pharmacol. 2007;71(3):644–53. doi:10.1124/mol.106.030635.

Kwon N-H, Kim J-S, Lee J-Y, Oh M-J, Choi D-C. DNA methylation and the expression of IL-4 and IFN-γ promoter genes in patients with bronchial asthma. J Clin Immunol. 2008;28(2):139–46. doi:10.1007/s10875-007-9148-1.

LaPlant Q, Vialou V, Covington HE, Dumitriu D, Feng J, Warren BL, et al. Dnmt3a regulates emotional behavior and spine plasticity in the nucleus accumbens. Nat Neurosci. 2010;13(9):1137–43. http://www.nature.com/neuro/journal/v13/n9/abs/nn.2619.html#supplementary-information.

Lee BH, Yegnasubramanian S, Lin X, Nelson WG. Procainamide is a specific inhibitor of DNA methyltransferase 1. J Biol Chem. 2005a;280(49):40749–56. doi:10.1074/jbc.M505593200.

Lee WJ, Shim J-Y, Zhu BT. Mechanisms for the inhibition of DNA methyltransferases by tea catechins and bioflavonoids. Mol Pharmacol. 2005b;68(4):1018–30. doi:10.1124/mol.104.008367.

Levenson JM, Roth TL, Lubin FD, Miller CA, Huang I-C, Desai P, et al. Evidence that DNA (cytosine-5) methyltransferase regulates synaptic plasticity in the Hippocampus. J Biol Chem. 2006;281(23):15763–73. doi:10.1074/jbc.M511767200.

Lewis DA, Hashimoto T, Volk DW. Cortical inhibitory neurons and schizophrenia. Nat Rev Neurosci. 2005;6(4):312–24.

Li Y, Tollefsbol TO. Impact on DNA methylation in cancer prevention and therapy by bioactive dietary components. Curr Med Chem. 2010;17(20):2141–51. doi:10.2174/092986710791299966.

Li LH, Olin EJ, Buskirk HH, Reineke LM. Cytotoxicity and mode of action of 5-azacytidine on L1210 leukemia. Cancer Res. 1970;30(11):2760–9.

Li C, Terashita Y, Tokoro M, Wakayama S, Wakayama T. Effect of DNA methyltransferase inhibitor, RG108, on in vitro development and ntES establishment rate in cloned mouse embryos. Reprod Fertil Dev. 2011;24:130. doi:10.1071/RDv24n1Ab36.

Li W, Yu M, Luo S, Liu H, Gao Y, Wilson JX, et al. DNA methyltransferase mediates dose-dependent stimulation of neural stem cell proliferation by folate. J Nutr Biochem. 2013;24(7):1295–301. http://dx.doi.org/10.1016/j.jnutbio.2012.11.001.

Li Y, Geng P, Jiang W, Wang Y, Yao J, Lin X, et al. Enhancement of radiosensitivity by 5-Aza-CdR through activation of G2/M checkpoint response and apoptosis in osteosarcoma cells. Tumor Biol. 2014;35(5):4831–9. doi:10.1007/s13277-014-1634-5.

Lin T-S, Lee H, Chen R-A, Ho M-L, Lin C-Y, Chen Y-H, et al. An association of DNMT3b protein expression with *P16INK4a* promoter hypermethylation in non-smoking female lung cancer with human papillomavirus infection. Cancer Lett. 2005;226(1):77–84. http://dx.doi.org/10.1016/j.canlet.2004.12.031.

Lin J, Haffner MC, Zhang Y, Lee BH, Brennen WN, Britton J, et al. Disulfiram is a DNA demethylating agent and inhibits prostate cancer cell growth. Prostate. 2011;71(4):333–43. doi:10.1002/pros.21247.

Lin N, Qin S, Luo S, Cui S, Huang G, Zhang X. Homocysteine induces cytotoxicity and proliferation inhibition in neural stem cells via DNA methylation *in vitro*. FEBS J. 2014;281(8):2088–96. doi:10.1111/febs.12764.

Liu Z, Xie Z, Jones W, Pavlovicz RE, Liu S, Yu J, et al. Curcumin is a potent DNA hypomethylation agent. Bioorg Med Chem Lett. 2009;19(3):706–9. http://dx.doi.org/10.1016/j.bmcl.2008.12.041.

Lloyd CM, Hawrylowicz CM. Regulatory T cells in asthma. Immunity. 2009;31(3):438–49. doi:10.1016/j.immuni.2009.08.007.

Logrip ML, Janak PH, Ron D. Escalating ethanol intake is associated with altered corticostriatal BDNF expression. J Neurochem. 2009;109(5):1459–68. doi:10.1111/j.1471-4159.2009.06073.x.

Lovinsky-Desir S, Ridder R, Torrone D, Maher C, Narula S, Scheuerman M, et al. DNA methylation of the allergy regulatory gene interferon gamma varies by age, sex, and tissue type in asthmatics. Clin Epigenet. 2014;6(1):9.

Lund G, Andersson L, Lauria M, Lindholm M, Fraga MF, Villar-Garea A, et al. DNA methylation polymorphisms precede any histological sign of atherosclerosis in mice lacking apolipoprotein E. J Biol Chem. 2004;279(28):29147–54. doi:10.1074/jbc.M403618200.

Luo S, Zhang X, Yu M, Yan H, Liu H, Wilson J, et al. Folic acid acts through DNA methyltransferases to induce the differentiation of neural stem cells into neurons. Cell Biochem Biophys. 2013;66(3):559–66. doi:10.1007/s12013-012-9503-6.

Luzzi A, Morettini F, Gazaneo S, Mundo L, Onnis A, Mannucci S, et al. HIV-1 Tat induces DNMT over-expression through microRNA dysregulation in HIV-related non Hodgkin lymphomas. Infect Agent Cancer. 2014;9(1):41. doi:10.1186/1750-9378-9-41.

Lyst MJ, Bird A. Rett syndrome: a complex disorder with simple roots. Nat Rev Genet. 2015;16(5):261–75. doi:10.1038/nrg3897.

Ma N, Chen W, Fan T, Tian Y, Zhang S, Zeng D, et al. Low temperature-induced DNA hypermethylation attenuates expression of *RhAG*, an *AGAMOUS* homolog, and increases petal number in rose (*Rosa hybrida*). BMC Plant Biol. 2015;15(1):237.

Machnes ZM, Huang TCT, Chang PKY, Gill R, Reist N, Dezsi G, et al. DNA methylation mediates persistent epileptiform activity *In vitro* and *In vivo*. PLoS One. 2013;8(10):e76299. doi:10.1371/journal.pone.0076299.

Mastroeni D, Grover A, Delvaux E, Whiteside C, Coleman PD, Rogers J. Epigenetic changes in Alzheimer's disease: decrements in DNA methylation. Neurobiol Aging. 2010;31(12):2025–37. http://dx.doi.org/10.1016/j.neurobiolaging.2008.12.005.

Matoušová M, Votruba I, Otmar M, Tloušťová E, Günterová J, Mertlíková-Kaiserová H. 2′-deoxy-5,6-dihydro-5-azacytidine—a less toxic alternative of 2′-deoxy-5-azacytidine: A comparative study of hypomethylating potential. Epigenetics. 2011;6(6):769–76. doi:10.4161/epi.6.6.16215.

Matrisciano F, Tueting P, Dalal I, Kadriu B, Grayson DR, Davis JM, et al. Epigenetic modifications of GABAergic interneurons are associated with the schizophrenia-like phenotype induced by

prenatal stress in mice. Neuropharmacology. 2013;68:184–94. http://dx.doi.org/10.1016/j.neuropharm.2012.04.013.

Meadows JP, Guzman-Karlsson MC, Phillips S, Holleman C, Posey JL, Day JJ, et al. DNA methylation regulates neuronal glutamatergic synaptic scaling. Sci Signal. 2015;8(382):ra61. doi:10.1126/scisignal.aab0715.

Medina-Franco JL, López-Vallejo F, Kuck D, Lyko F. Natural products as DNA methyltransferase inhibitors: a computer-aided discovery approach. Mol Divers. 2011;15(2):293–304. doi:10.1007/s11030-010-9262-5.

Melas PA, Rogdaki M, Lennartsson A, Björk K, Qi H, Witasp A, et al. Antidepressant treatment is associated with epigenetic alterations in the promoter of P11 in a genetic model of depression. Int J Neuropsychopharmacol. 2012;15(5):669–79. doi:10.1017/s1461145711000940.

Miao C-g, Huang C, Huang Y, Yang Y-y, He X, Zhang L, et al. MeCP2 modulates the canonical Wnt pathway activation by targeting SFRP4 in rheumatoid arthritis fibroblast-like synoviocytes in rats. Cell Signal. 2013;25(3):598–608. http://dx.doi.org/10.1016/j.cellsig.2012.11.023.

Mikovits JA, Young HA, Vertino P, Issa J-PJ, Pitha PM, Turcoski-Corrales S, et al. Infection with human immunodeficiency virus type 1 upregulates DNA methyltransferase, resulting in de novo methylation of the gamma interferon (IFN-γ) promoter and subsequent downregulation of IFN-γ production. Mol Cell Biol. 1998;18(9):5166–77. doi:10.1128/mcb.18.9.5166.

Miller CA, Sweatt JD. Covalent modification of DNA regulates memory formation. Neuron. 2007;53(6):857–69. http://dx.doi.org/10.1016/j.neuron.2007.02.022.

Miller CA, Gavin CF, White JA, Parrish RR, Honasoge A, Yancey CR, et al. Cortical DNA methylation maintains remote memory. Nat Neurosci. 2010;13(6):664–6. http://www.nature.com/neuro/journal/v13/n6/abs/nn.2560.html#supplementary-information.

Mitchnick KA, Creighton S, O'Hara M, Kalisch BE, Winters BD. Differential contributions of de novo and maintenance DNA methyltransferases to object memory processing in the rat hippocampus and perirhinal cortex – a double dissociation. Eur J Neurosci. 2015;41(6):773–86. doi:10.1111/ejn.12819.

Momparler RL, Derse D. Kinetics of phosphorylation of 5-aza-2′-deoxyycytidine by deoxycytidine kinase. Biochem Pharmacol. 1979;28(8):1443–4.

Nature Editorials. Suicide watch. Nature. 2014;506:131.

Nelson ED, Kavalali ET, Monteggia LM. Activity-dependent suppression of miniature neurotransmission through the regulation of DNA methylation. J Neurosci. 2008;28(2):395–406. doi:10.1523/jneurosci.3796-07.2008.

Nestor CE, Barrenäs F, Wang H, Lentini A, Zhang H, Bruhn S, et al. DNA methylation changes separate allergic patients from healthy controls and may reflect altered CD4+ T-cell population structure. PLoS Genet. 2014;10(1):e1004059. doi:10.1371/journal.pgen.1004059.

Noh JS, Sharma RP, Veldic M, Salvacion AA, Jia X, Chen Y, et al. DNA methyltransferase 1 regulates reelin mRNA expression in mouse primary cortical cultures. Proc Natl Acad Sci U S A. 2005;102(5):1749–54. doi:10.1073/pnas.0409648102.

Notari RE, Deyoung JL. Kinetics and mechanisms of degradation of the antileukemic agent 5-azacytidine in aqueous solutions. J Pharm Sci. 1975;64(7):1148–57. doi:10.1002/jps.2600640704.

Nutt DJ, King LA, Phillips LD. Drug harms in the UK: a multicriteria decision analysis. Lancet. 2010;376(9752):1558–65. doi:10.1016/s0140-6736(10)61462-6.

Oh YS, Jeong S-G, Cho G-W. Anti-senescence effects of DNA methyltransferase inhibitor RG108 in human bone marrow mesenchymal stromal cells. Biotechnol Appl Biochem. 2015a:[Epub ahead of print]. http://dx.doi.org/10.1002/bab.393. doi:10.1002/bab.1393.

Oh YS, Kim SH, Cho G-W. Functional restoration of amyotrophic lateral sclerosis patient-derived mesenchymal stromal cells through inhibition of DNA methyltransferase. Cell Mol Neurobiol. 2015b. [Epub ahead of print] http://dx.doi.org/10.1007/s10571-015-0242-2. doi:10.1007/s10571-015-0242-2.

Orta ML, Höglund A, Calderón-Montaño JM, Domínguez I, Burgos-Morón E, Visnes T, et al. The PARP inhibitor Olaparib disrupts base excision repair of 5-aza-2′-deoxycytidine lesions. Nucleic Acids Res. 2014;42(14):9108–20. doi:10.1093/nar/gku638.

Park J, Jang KL. Hepatitis C virus represses E-cadherin expression via DNA methylation to induce epithelial to mesenchymal transition in human hepatocytes. Biochem Biophys Res Commun. 2014;446(2):561–7. http://dx.doi.org/10.1016/j.bbrc.2014.03.009.

Parrish RR, Albertson AJ, Buckingham SC, Hablitz JJ, Mascia KL, Haselden WD, et al. status epilepticus triggers early and late alterations in brain-derived neurotrophic factor and NMDA glutamate receptor GRIN2B DNA methylation levels in the Hippocampus. Neuroscience. 2013;0:602–19. doi:10.1016/j.neuroscience.2013.06.029.

Paschos K, Smith P, Anderton E, Middeldorp JM, White RE, Allday MJ. Epstein-Barr virus latency in B cells leads to epigenetic repression and CpG methylation of the tumour suppressor gene Bim. PLoS Pathog. 2009;5(6):e1000492. doi:10.1371/journal.ppat.1000492.

Peat JR, Reik W. Incomplete methylation reprogramming in SCNT embryos. Nat Genet. 2012;44(9):965–6.

Penter L, Maier B, Frede U, Hackner B, Carell T, Hagemeier C, et al. A rapid screening system evaluates novel inhibitors of DNA methylation and suggests F-box proteins as potential therapeutic targets for high-risk neuroblastoma. Targeted Oncol. 2015. [Epub ahead of print]. doi:10.1007/s11523-014-0354-5.

Pereira R, Benedetti R, Pérez-Rodriguez S, Nebbioso A, GarcÃia-Rodriguez J, Carafa V, et al. Indole-Derived Psammaplin A analogues as epigenetic modulators with multiple inhibitory activities. J Med Chem. 2012;55(22):9467–91.

Philibert RA, Plume JM, Gibbons FX, Brody GH, Beach SRH. The impact of recent alcohol use on genome wide DNA methylation signatures. Front Genet. 2012;3:54. doi:10.3389/fgene.2012.00054.

Ponomarev I, Wang S, Zhang L, Harris RA, Mayfield RD. Gene coexpression networks in human brain identify epigenetic modifications in alcohol dependence. J Neurosci. 2012;32(5):1884–97. doi:10.1523/jneurosci.3136-11.2012.

Raffoux E, Cras A, Recher C, Boëlle P-Y, de Labarthe A, Turlure P, et al. Phase 2 clinical trial of 5-azacitidine, valproic acid, and all-trans retinoic acid in patients with high-risk acute myeloid leukemia or myelodysplastic syndrome. Oncotarget. 2010;1(1):34–42.

Rilova E, Erdmann A, Gros C, Masson V, Aussagues Y, Poughon-Cassabois V, et al. Design, synthesis and biological evaluation of 4-amino-N-(4-aminophenyl)benzamide analogues of quinoline-based SGI-1027 as inhibitors of DNA methylation. ChemMedChem. 2014;9(3):590–601. doi:10.1002/cmdc.201300420.

Ronald PC. The molecular basis of disease resistance in rice. Plant Mol Biol. 1997;35(1–2):179–86. doi:10.1023/a:1005750811983.

Roth TL, Lubin FD, Funk AJ, Sweatt JD. Lasting epigenetic influence of early-life adversity on the BDNF gene. Biol Psychiatry. 2009;65(9):760–9. doi:10.1016/j.biopsych.2008.11.028.

Runyon RS, Cachola LM, Rajeshuni N, Hunter T, Garcia M, Ahn R, et al. Asthma discordance in twins is linked to epigenetic modifications of T cells. PLoS One. 2012;7(11):e48796. doi:10.1371/journal.pone.0048796.

Saggioro D, Forino M, Chieco-Bianchi L. Transcriptional block of HTLV-I LTR by sequence-specific methylation. Virology. 1991;182(1):68–75. http://dx.doi.org/10.1016/0042-6822(91)90649-V.

Saini M, Selokar NL, Agrawal H, Singla SK, Chauhan MS, Manik RS, et al. Treatment of buffalo (*Bubalus bubalis*) donor cells with trichostatin A and 5-aza-2′-deoxycytidine alters their growth characteristics, gene expression and epigenetic status and improves the in vitro developmental competence, quality and epigenetic status of cloned embryos. Reprod Fertil Dev. 2014. [Epub ahead of print]. doi:10.1071/RD14176.

Sales AJ, Biojone C, Terceti MS, Guimarães FS, Gomes MVM, Joca SRL. Antidepressant-like effect induced by systemic and intra-hippocampal administration of DNA methylation inhibitors. Br J Pharmacol. 2011;164(6):1711–21. doi:10.1111/j.1476-5381.2011.01489.x.

Sano H, Kamada I, Youssefian S, Katsumi M, Wabiko H. A single treatment of rice seedlings with 5-azacytidine induces heritable dwarfism and undermethylation of genomic DNA. Mol Gen Genet. 1990;220(3):441–7. doi:10.1007/bf00391751.

Santi DV, Norment A, Garrett CE. Covalent bond formation between a DNA-cytosine methyltransferase and DNA containing 5-azacytosine. Proc Natl Acad Sci U S A. 1984;81(22):6993–7.

Satta R, Maloku E, Zhubi A, Pibiri F, Hajos M, Costa E, et al. Nicotine decreases DNA methyltransferase 1 expression and glutamic acid decarboxylase 67 promoter methylation in GABAergic interneurons. Proc Natl Acad Sci U S A. 2008;105(42):16356–61. doi:10.1073/pnas.0808699105.

Savickiene J, Treigyte G, Jonusiene V, Bruzaite R, Borutinskaite V-V, Navakauskiene R. Epigenetic changes by zebularine leading to enhanced differentiation of human promyelocytic leukemia NB4 and KG1 cells. Mol Cell Biochem. 2012;359(1–2):245–61. doi:10.1007/s11010-011-1019-7.

Scarpa S, Fuso A, D'Anselmi F, Cavallaro RA. Presenilin 1 gene silencing by *S*-adenosylmethionine: a treatment for Alzheimer disease?. FEBS Lett. 2003;541(1–3):145–8. http://dx.doi.org/10.1016/S0014-5793(03)00277-1.

Schiano C, Vietri MT, Grimaldi V, Picascia A, De Pascale MR, Napoli C. Epigenetic-related therapeutic challenges in cardiovascular disease. Trends Pharmacol Sci. 2015;36(4):226–35. http://dx.doi.org/10.1016/j.tips.2015.02.005.

Schubert HL, Blumenthal RM, Cheng X. Many paths to methyltransfer: a chronicle of convergence. Trends Biochem Sci. 2003;28(6):329–35. doi:10.1016/s0968-0004(03)00090-2.

Shahid SK, Kharitonov SA, Wilson NM, Bush A, Barnes PJ. Increased interleukin-4 and decreased interferon-γ in exhaled breath condensate of children with asthma. Am J Respir Crit Care Med. 2002;165(9):1290–3. doi:10.1164/rccm.2108082.

Sheikhnejad G, Brank A, Christman JK, Goddard A, Alvarez E, Ford Jr H, et al. Mechanism of inhibition of DNA (cytosine C5)-methyltransferases by oligodeoxyribonucleotides containing 5,6-dihydro-5-azacytosine. J Mol Biol. 1999;285(5):2021–34. doi:10.1006/jmbi.1998.2426.

Siedlecki P, Boy RG, Musch T, Brueckner B, Suhai S, Lyko F, et al. Discovery of two novel, small-molecule inhibitors of DNA methylation. J Med Chem. 2006;49(2):678–83. doi:10.1021/jm050844z.

Silverman LR, Verma A, Odchimar-Reissig R, Leblanc A, Najfeld V, Gabrilove J, et al. A phase I trial of epigenetic modulators vorinostat, in combination with azacitidine (azaC) in patients with the myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML): a study of the New York Cancer Consortium [Abstract]. Blood (ASH Annual Meeting Abstracts). 2008;112:3656.

Silverman LR, Holland J, Sandvold ML, Myhren F, Eriksen OH, inventors; Analogues d'azacytidine et leurs utilisations. 2009.

Singh N, Dueñas-González A, Lyko F, Medina-Franco JL. Molecular modeling and molecular dynamics studies of hydralazine with human DNA methyltransferase 1. ChemMedChem. 2009a;4(5):792–9. doi:10.1002/cmdc.200900017.

Singh RP, Shiue K, Schomberg D, Zhou FC. Cellular epigenetic modifications of neural stem cell differentiation. Cell Transplant. 2009b;18(10):1197–211. doi:10.3727/096368909x12483162197204.

Smith AK, Conneely KN, Kilaru V, Mercer KB, Weiss TE, Bradley B, et al. Differential immune system DNA methylation and cytokine regulation in post-traumatic stress disorder. Am J Med Genet B Neuropsychiatr Genet. 2011;156(6):700–8. doi:10.1002/ajmg.b.31212.

Song Y, Zhang C. Hydralazine inhibits human cervical cancer cell growth in vitro in association with APC demethylation and re-expression. Cancer Chemother Pharmacol. 2009;63(4):605–13. doi:10.1007/s00280-008-0773-z.

Sorm F, Piskala A, Cihak A, Vesely J. 5-Azacytidine, a new, highly effective cancerostatic. Experientia. 1964;20(4):202–3.

Stresemann C, Brueckner B, Musch T, Stopper H, Lyko F. Functional diversity of DNA methyltransferase inhibitors in human cancer cell lines. Cancer Res. 2006;66(5):2794–800. doi:10.1158/0008-5472.can-05-2821.

Struck A-W, Thompson ML, Wong LS, Micklefield J. *S*-adenosyl-methionine-dependent methyltransferases: highly versatile enzymes in biocatalysis. Biosynthesis and other biotechnological applications. Chembiochem. 2012;13(18):2642–55.

Suh N, Pezzuto JM. Strawberry fields forever? Cancer Prev Res. 2012;5(1):30–3. doi:10.1158/1940-6207.capr-11-0534.

Sun J, Awakawa T, Noguchi H, Abe I. Induced production of mycotoxins in an endophytic fungus from the medicinal plant *Datura stramonium L.* Bioorg Med Chem Lett. 2012;22(20):6397–400. http://dx.doi.org/10.1016/j.bmcl.2012.08.063.

Sun H, Kennedy PJ, Nestler EJ. Epigenetics of the depressed brain: role of histone acetylation and methylation. Neuropsychopharmacology. 2013;38(1):124–37.

Sutcliffe JS, Nelson DL, Zhang F, Pieretti M, Caskey CT, Saxe D, et al. DNA methylation represses FMR-1 transcription in fragile X syndrome. Hum Mol Gen. 1992;1(6):397–400. doi:10.1093/hmg/1.6.397.

Suzuki T, Tanaka R, Hamada S, Nakagawa H, Miyata N. Design, synthesis, inhibitory activity, and binding mode study of novel DNA methyltransferase 1 inhibitors. Bioorg Med Chem Lett. 2010;20(3):1124–7.

Szyf M. Prospects for the development of epigenetic drugs for CNS conditions. Nat Rev Drug Discov. 2015;14(7):461–74. doi:10.1038/nrd4580.

Tabolacci E, Pietrobono R, Moscato U, Oostra BA, Chiurazzi P, Neri G. Differential epigenetic modifications in the FMR1 gene of the fragile X syndrome after reactivating pharmacological treatments. Eur J Hum Genet. 2005;13(5):641–8.

Taylor SM, Jones PA. Multiple new phenotypes induced in 10T12 and 3T3 cells treated with 5-azacytidine. Cell. 1979;17(4):771–9. http://dx.doi.org/10.1016/0092-8674(79)90317-9.

Thangavel J, Malik AB, Elias HK, Rajasingh S, Simpson AD, Sundivakkam PK, et al. Combinatorial therapy with acetylation and methylation modifiers attenuates lung vascular hyperpermeability in endotoxemia-induced mouse inflammatory lung injury. Am J Pathol. 2014;184(8):2237–49. doi:10.1016/j.ajpath.2014.05.008.

Thangavel J, Samanta S, Rajasingh S, Barani B, Xuan Y-T, Dawn B, et al. Epigenetic modifiers reduce inflammation and modulate macrophage phenotype during endotoxemia-induced acute lung injury. J Cell Sci. 2015;128(16):3094–105. doi:10.1242/jcs.170258.

Tolg C, Sabha N, Cortese R, Panchal T, Ahsan A, Soliman A, et al. Uropathogenic E. coli infection provokes epigenetic downregulation of CDKN2A (p16INK4A) in uroepithelial cells. Lab Invest. 2011;91(6):825–36.

Tremolizzo L, Doueiri M-S, Dong E, Grayson DR, Davis J, Pinna G, et al. Valproate corrects the schizophrenia-like epigenetic behavioral modifications induced by methionine in mice. Biol Psychiatry. 2005;57(5):500–9. http://dx.doi.org/10.1016/j.biopsych.2004.11.046.

Uchida S, Hara K, Kobayashi A, Otsuki K, Yamagata H, Hobara T, et al. Epigenetic status of *Gdnf* in the ventral striatum determines susceptibility and adaptation to daily stressful events. Neuron. 2011;69(2):359–72. http://dx.doi.org/10.1016/j.neuron.2010.12.023.

Uddin M, Aiello AE, Wildman DE, Koenen KC, Pawelec G, de los Santos R, et al. Epigenetic and immune function profiles associated with posttraumatic stress disorder. Proc Natl Acad Sci U S A. 2010;107(20):9470–5. doi:10.1073/pnas.0910794107.

Valente S, Liu Y, Schnekenburger M, Zwergel C, Cosconati S, Gros C, et al. Selective non-nucleoside inhibitors of human DNA methyltransferases active in cancer including in cancer stem cells. J Med Chem. 2014;57(3):701–13. doi:10.1021/jm4012627.

van Bemmel DM, Brank AS, Eritja R, Marquez VE, Christman JK. DNA (Cytosine-C5) methyltransferase inhibition by oligodeoxyribonucleotides containing 2-(1H)-pyrimidinone (zebularine aglycon) at the enzymatic target site. Biochem Pharmacol. 2009;78(6):633–41. doi:10.1016/j.bcp.2009.05.017.

Van Lint C, Bouchat S, Marcello A. HIV-1 transcription and latency: an update. Retrovirology. 2013;10(1):67. doi:10.1186/1742-4690-10-67.

Veldic M, Caruncho HJ, Liu WS, Davis J, Satta R, Grayson DR, et al. DNA-methyltransferase 1 mRNA is selectively overexpressed in telencephalic GABAergic interneurons of schizophrenia brains. Proc Natl Acad Sci U S A. 2004;101(1):348–53. doi:10.1073/pnas.2637013100.

Villar-Garea A, Fraga MF, Espada J, Esteller M. Procaine is a DNA-demethylating agent with growth-inhibitory effects in human cancer cells. Cancer Res. 2003;63(16):4984–9.

Vyskot B, Koukalová B, Kovařík A, Sachambula L, Reynolds D, Bezděk M. Meiotic transmission of a hypomethylated repetitive DNA family in tobacco. Theor Appl Genetics. 1995;91(4):659–64. doi:10.1007/bf00223294.

Wang X, Chao L, Jin G, Ma G, Zang Y, Sun J. Association between CpG island methylation of the WWOX gene and its expression in breast cancers. Tumor Biol. 2009;30(1):8–14.

Wang X, Sena Filho JG, Hoover AR, King JB, Ellis TK, Powell DR, et al. Chemical epigenetics alters the secondary metabolite composition of guttate excreted by an Atlantic-forest-soil-derived Penicillium citreonigrum. J Nat Prod. 2010a;73(5):942–8. doi:10.1021/np100142h.
Wang X, Zhu H, Snieder H, Su S, Munn D, Harshfield G, et al. Obesity related methylation changes in DNA of peripheral blood leukocytes. BMC Med. 2010b;8:87. doi:10.1186/1741-7015-8-87.
Wang L, Zhang Y, Li R, Chen Y, Pan X, Li G, et al. 5-aza-2′-deoxycytidine enhances the radiosensitivity of breast cancer cells. Cancer Biother Radiopharm. 2013a;28(1):34–44. doi:10.1089/cbr.2012.1170.
Wang Y, Wang X, Li R, Yang Z-F, Wang Y-Z, Gong X-L, et al. A DNA methyltransferase inhibitor, 5-Aza-2′-deoxycytidine, exacerbates neurotoxicity and upregulates Parkinson's disease-related genes in dopaminergic neurons. CNS Neurosci Ther. 2013b;19(3):183–90. doi:10.1111/cns.12059.
Warnault V, Darcq E, Levine A, Barak S, Ron D. Chromatin remodeling [mdash] a novel strategy to control excessive alcohol drinking. Transl Psychiatry. 2013;3:e231. doi:10.1038/tp.2013.4.
Watanabe A, Yamada Y, Yamanaka S. Epigenetic regulation in pluripotent stem cells: a key to breaking the epigenetic barrier. Philos Trans R Soc B Biol Sci. 2013;368(1609):20120292. doi:10.1098/rstb.2012.0292.
Watson CJ, Collier P, Tea I, Neary R, Watson JA, Robinson C, et al. Hypoxia-induced epigenetic modifications are associated with cardiac tissue fibrosis and the development of a myofibroblast-like phenotype. Hum Mol Gen. 2014;23(8):2176–88. doi:10.1093/hmg/ddt614.
White GP, Hollams EM, Yerkovich ST, Bosco A, Holt BJ, Bassami MR, et al. CpG methylation patterns in the IFNγ promoter in naive T cells: Variations during Th1 and Th2 differentiation and between atopics and non-atopics. Pediatr Allergy Immunol. 2006;17(8):557–64. doi:10.1111/j.1399-3038.2006.00465.x.
Winders BR, Schwartz RH, Bruniquel D. A distinct region of the murine IFN-γ promoter is hypomethylated from early T cell development through mature naive and Th1 cell differentiation, but is hypermethylated in Th2 cells. J Immunol. 2004;173(12):7377–84. doi:10.4049/jimmunol.173.12.7377.
Xiang L, Dong W, Wang R, Wei J, Qiu G, Cen J, et al. All-trans retinoic acid enhances the effect of 5-aza-2′-deoxycytidine on p16INK4a demethylation, and the two drugs synergistically activate retinoic acid receptor β gene expression in the human erythroleukemia K562 cell line. Oncol Lett. 2014;8(1):117–22. doi:10.3892/ol.2014.2133.
Xiao D, Dasgupta C, Chen M, Zhang K, Buchholz J, Xu Z, et al. Inhibition of DNA methylation reverses norepinephrine-induced cardiac hypertrophy in rats. Cardiovasc Res. 2014;101(3):373–82. doi:10.1093/cvr/cvt264.
Xu W, Li Z, Yu B, He X, Shi J, Zhou R, et al. Effects of DNMT1 and HDAC inhibitors on gene-specific methylation reprogramming during porcine somatic cell nuclear transfer. PLoS One. 2013;8(5):e64705. doi:10.1371/journal.pone.0064705.
Yamauchi T, Kamon J, Minokoshi Y, Ito Y, Waki H, Uchida S, et al. Adiponectin stimulates glucose utilization and fatty-acid oxidation by activating AMP-activated protein kinase. Nat Med. 2002;8(11):1288–95. http://www.nature.com/nm/journal/v8/n11/suppinfo/nm788_S1.html.
Yang H, Hoshino K, Sanchez-Gonzalez B, Kantarjian H, Garcia-Manero G. Antileukemia activity of the combination of 5-aza-2′-deoxycytidine with valproic acid. Leukemia Res. 2005;29(7):739–48. http://dx.doi.org/10.1016/j.leukres.2004.11.022.
Yang D, Torres CM, Bardhan K, Zimmerman M, McGaha TL, Liu K. Decitabine and vorinostat cooperate to sensitize colon carcinoma cells to Fas ligand-induced apoptosis *In vitro* and tumor suppression *in vivo*. J Immunol. 2012;188(9):4441–9. doi:10.4049/jimmunol.1103035.
Yang P-M, Lin Y-T, Shun C-T, Lin S-H, Wei T-T, Chuang S-H, et al. Zebularine inhibits tumorigenesis and stemness of colorectal cancer via p53-dependent endoplasmic reticulum stress. Sci Rep. 2013;3:3219. doi:10.1038/srep03219, http://www.nature.com/articles/srep03219#supplementary-information.
Yie Y, Zhao S, Tang Q, Zheng F, Wu J, Yang L, et al. Ursolic acid inhibited growth of hepatocellular carcinoma HepG2 cells through AMPKalpha-mediated reduction of DNA methyltransferase 1. Mol Cell Biochem. 2015;402(1–2):63–74. doi:10.1007/s11010-014-2314-x.

Yin L, Chung WO. Epigenetic regulation of human β-defensin 2 and CC chemokine ligand 20 expression in gingival epithelial cells in response to oral bacteria. Mucosal Immunol. 2011;4(4):409–19. doi:10.1038/mi.2010.83.

Yin L-J, Zhang Y, Lv P-P, He W-H, Wu Y-T, Liu A-X, et al. Insufficient maintenance DNA methylation is associated with abnormal embryonic development. BMC Med. 2012;10(1):26. doi:10.1186/1741-7015-10-26.

Yogelzang NJ, Herndon 2nd JE, Cirrincione C, Harmon DC, Antman KH, Corson JM, et al. Dihydro-5-azacytidine in malignant mesothelioma. A phase II trial demonstrating activity accompanied by cardiac toxicity. Cancer and Leukemia Group B. Cancer. 1997;79(11): 2237–42.

Yoo CB, Jeong S, Egger G, Liang G, Phiasivongsa P, Tang C, et al. Delivery of 5-aza-2′-deoxycytidine to cells using oligodeoxynucleotides. Cancer Res. 2007;67(13):6400–8. doi:10.1158/0008-5472.CAN-07-0251.

Yoo J, Choi S, Medina-Franco JL. Molecular modeling studies of the novel inhibitors of DNA methyltransferases SGI-1027 and CBC12: implications for the mechanism of inhibition of DNMTs. PLoS One. 2013;8(4):e62152. doi:10.1371/journal.pone.0062152.

Yu Q, Zhou B, Zhang Y, Nguyen ET, Du J, Glosson NL, et al. DNA methyltransferase 3a limits the expression of interleukin-13 in T helper 2 cells and allergic airway inflammation. Proc Natl Acad Sci U S A. 2012;109(2):541–6. doi:10.1073/pnas.1103803109.

Zaina S, Heyn H, Carmona FJ, Varol N, Sayols S, Condom E, et al. DNA methylation Map of human atherosclerosis. Circ Cardiovasc Genet. 2014;7(5):692–700. doi:10.1161/circgenetics.113.000441.

Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SWL, Chen H, et al. Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. Cell. 2006;126(6):1189–201. http://dx.doi.org/10.1016/j.cell.2006.08.003.

Zhao Q, Fan J, Hong W, Li L, Wu M. Inhibition of cancer cell proliferation by 5-fluoro-2′-deoxycytidine, a DNA methylation inhibitor, through activation of DNA damage response pathway. Springerplus. 2012;1(1):65.

Zhao S, Wu J, Zheng F, Tang Q, Yang L, Li L, et al. β-elemene inhibited expression of DNA methyltransferase 1 through activation of ERK1/2 and AMPKalpha signalling pathways in human lung cancer cells: the role of Sp1. J Cell Mol Med. 2015;19(3):630–41. doi:10.1111/jcmm.12476.

Zhubi A, Veldic M, Puri NV, Kadriu B, Caruncho H, Loza I, et al. An upregulation of DNA-methyltransferase 1 and 3a expressed in telencephalic GABAergic neurons of schizophrenia patients is also detected in peripheral blood lymphocytes. Schizophr Res. 2009;111(1–3):115–22. http://dx.doi.org/10.1016/j.schres.2009.03.020.

Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. Nat Genet. 2007;39(1):61–9. http://www.nature.com/ng/journal/v39/n1/suppinfo/ng1929_S1.html.

Zovkic IB, Sweatt JD. Epigenetic mechanisms in learned fear: implications for PTSD. Neuropsychopharmacol Rev. 2013;38(1):77–93.

Rewriting DNA Methylation Signatures at Will: The *Curable Genome* Within Reach?

Sabine Stolzenburg, Désirée Goubert, and Marianne G. Rots

Abstract

Epigenetic regulation of gene expression is vital for the maintenance of genome integrity and cell phenotype. In addition, many different diseases have underlying epigenetic mutations, and understanding their role and function may unravel new insights for diagnosis, treatment, and even prevention of diseases. It was an important breakthrough when epigenetic alterations could be gene-specifically manipulated using epigenetic regulatory proteins in an approach termed epigenetic editing. Epigenetic editors can be designed for virtually any gene by targeting effector domains to a preferred sequence, where they write or erase the desired epigenetic modification. This chapter describes the tools for editing DNA methylation signatures and their applications. In addition, we explain how to achieve targeted DNA (de)methylation and discuss the advantages and disadvantages of this approach. Silencing genes directly at the DNA methylation level instead of targeting the protein and/or RNA is a major improvement, as repression is achieved at the source of expression, potentially eliminating the need for continuous administration. Re-expression of silenced genes by targeted demethylation might closely represent the natural situation, in which all transcript variants might be expressed in a sustainable manner. Altogether epigenetic editing, for example, by rewriting DNA methylation, will assist in realizing the *curable genome* concept.

S. Stolzenburg • D. Goubert • M.G. Rots (✉)
Epigenetic Editing Research Group, Department of Pathology and Medical Biology, University Medical Center Groningen, University of Groningen, Hanzeplein 1 EA11, Groningen 9713 GZ, The Netherlands
e-mail: m.g.rots@umcg.nl

© Springer International Publishing Switzerland 2016

A. Jeltsch, R.Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*, Advances in Experimental Medicine and Biology 945,
DOI 10.1007/978-3-319-43624-1_17

Abbreviations

ATF	Artificial transcription factor
ChIP	Chromatin immunoprecipitation
CpG	Cytosine–phosphate–guanine
CRISPRs	Clustered regulatory interspaced palindromic repeats
DNMT	DNA methyltransferase
ncRNA	Nonprotein-coding RNA
sgRNA	Single-guide RNA
TALEs	Transcription activator-like effectors
TDG	Thymidine–DNA glycosylase
TET	Ten–eleven translocation
ZF	Zinc finger

1 Introduction

The term epigenetics was coined by Conrad Waddington back in 1942, who defined epigenetics as the branch of biology that studies the causal interactions between genes and their products which bring the phenotype into being (Waddington 2012). This definition has evolved over time to the current understanding of epigenetics referring to the study of heritable changes in gene expression that occur independent of changes in the primary DNA sequence (Sharma et al. 2010). The basic unit of chromatin comprises the nucleosome, which consists of approximately 146 base pairs (bps) of DNA wrapped around an octamer containing two copies of each of the core histones H2A, H2B, H3, and H4. Biochemical modifications on DNA and histones, as well as the nuclear context, influence the three-dimensional structure of chromatin. The main covalent chemical modification on DNA itself is the methylation of cytosines at sites where it is followed by a guanine base (CpGs). Additionally, posttranslational histone modification (PTMs), nucleosome positioning, and the expression of nonprotein-coding RNAs (ncRNAs) are important epigenetic modifications.

A huge number of data has been generated on how epigenetics regulate gene expression; however, the majority of these data are only correlative in nature. In order to study the causative role of a particular epigenetic modification at a given genomic site, epigenetic editing approaches have been exploited in the recent years (de Groote et al. 2012; Jurkowski et al. 2015). Epigenetic editing refers to the technology of actively rewriting epigenetic signatures at a genomic locus of interest. Toward this end, molecular tools – mostly developed and used in genome engineering (Gaj et al. 2013) – have been employed that allow DNA binding at a predefined genomic locus (Kungulovski and Jeltsch 2015). The most frequently used devices for gene targeting are self-engineered zinc finger (ZF) proteins, transcription activator-like effectors (TALEs), or the recently introduced clustered regulatory interspaced palindromic repeats (CRISPRs) system, which is based on DNA targeting by RNA molecules, the so-called single-guide RNAs (sgRNAs) (Fig. 1).

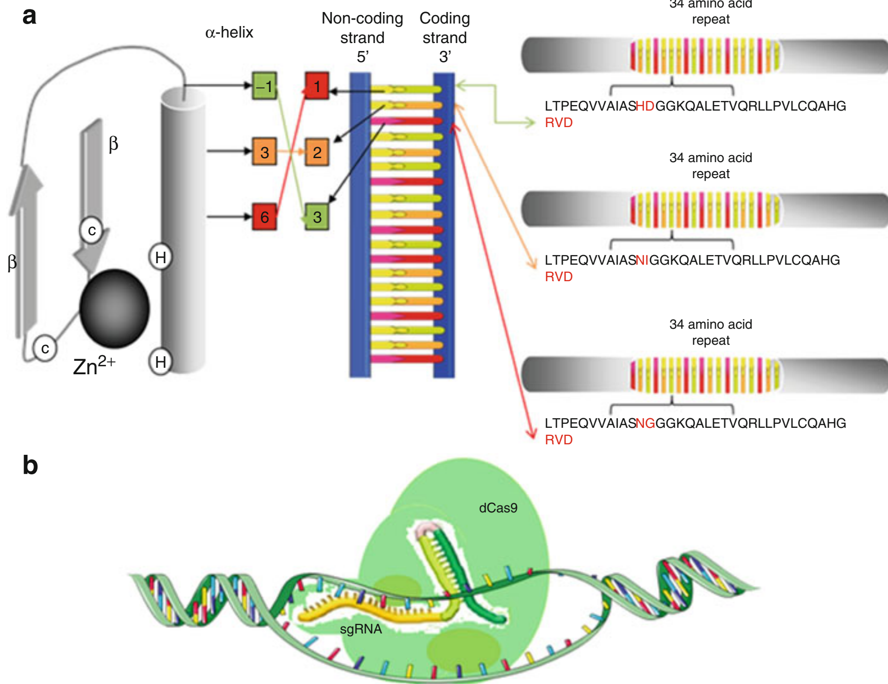


Fig. 1 Schematic representation of the DNA-binding domains of the most commonly used molecular tools in epigenetic editing. (**a**) The ZF protein on the left consists of approximately 30 amino acids, in which AAs at positions -1, 3, and 6 in the alpha-helix of the ZF protein recognize the third, second, and first base pair of the 5'–3' target sequence. Specificity can be increased by linking several ZFs together. The TALEs on the right consist of different monomers of approximately 34 AAs that are variable at positions 12 and 13 (=RVD), which are responsible for targeting a specific base pair within the DNA sequence. Notice that in comparison to the ZF, three times as much AAs are responsible for targeting the same amount of base pairs. (**b**) sgRNAs guide the CRISPR–dCas9 system to a particular sequence of approximately 20 base pairs. *ZF* zinc finger, *AA* amino acid, *TALE* transcription activator-like effector, *RVD* repeat variable di-residue, *sgRNA* single-guide RNA, *CRISPR* clustered regulatory interspaced palindromic repeat

ZF proteins are naturally occurring transcription factors forming the largest group of all transcription factors in the human genome (Vaquerizas et al. 2009). They consist of approximately 30 amino acids, wherein a stretch of seven amino acids is responsible for the recognition of 3–4 bps in the major groove of double-stranded DNA. During ZF binding, the amino acids at positions -1, 3, and 6 in the alpha-helix of the ZF protein recognize the third, second, and first base pair of the 5'–3' target sequence (Fig. 1a, left side). In 1996, Kim and Berg published the crystal structure of a designed ZF protein, which led to a refinement of this code, because it revealed an additional bond between a certain amino acid at position 2 in the ZF alpha-helix and the 4th base in the antisense strand of the DNA, which is at the same time the complement nucleotide of the 2 triplicate, recognized by amino acid 6 of the second ZF protein (Kim and Berg 1996).

Based on this knowledge, different DNA sequences can be targeted by engineering ZF proteins via exchanging amino acids of the α-helix to bind three base pairs of choice (Vandevenne et al. 2013). The specificity of ZF proteins is subsequently increased by linking several ZF domains together, so that, for example, a six-finger ZF protein will recognize 18 base pairs of target DNA (Fig. 1a, left side). When generating these ZF arrays, the selection procedure strongly determines the potency, and target site overlap or cross talk may complicate the array generation (Mussolino et al. 2011).

In 1994, Klug and colleagues engineered the first ZF protein successfully targeting and repressing the *BCR–ABL* fusion oncogene (Choo et al. 1994). Since this pioneering work was conducted, engineered ZF proteins have been used in fusion with nucleases (*molecular scissors*) or transcriptional activators and repressors (*artificial transcription factors*, ATFs) to target a multitude of endogenous genes (de Groote et al. 2012). The relatively small size and low immunogenicity of ZF proteins are a major advantage compared to other DNA-targeting proteins (Falahi et al. 2015; Mussolino et al. 2011). Importantly, the potential of ZF proteins as molecular scissors for therapeutic applications is explored in clinical trials (Ledford 2011).

TALEs are derived from plant pathogenic bacteria where they are used to modulate host gene expression (Boch and Bonas 2010). Upon injection into the plant cells, TALEs are imported into the nucleus where they bind specific sequences of the host cell genome and activate transcription. Like ZFs, TALEs also consist of individual modules (Jurkowski et al. 2015): each monomer (of 33 or 34 amino acids) differs at amino acid positions 12 and 13, a region called repeat variable diresidue (RVD) (Fig. 1a, right side). These hypervariable residues mediate binding to the target DNA site. Each RVD recognizes one nucleotide within the DNA-binding site (HD=C, NI=A, NG=T, NN=G), allowing for a straightforward design. Subsequently, transcriptional activators, repressors, or nucleases can be fused to the TALE DNA-binding domain (DBD) for targeted gene expression modification. Targeting efficiencies of the TALE DBD range from 25 to 95 % (Miller et al. 2011; Maeder et al. 2013b), and new assembly methods are now available to improve the generation of more efficient TALEs (Reyon et al. 2012; Briggs et al. 2012). Considerable progress has been made in the design, development, and characterization of TALEs (Cermak et al. 2011).

Another breakthrough technology was introduced early in 2013 the flexible CRISPRs–CRISPR-associated proteins (CRISPR–Cas) system, which revolutionized biomedical research because of its ease, low cost, and flexibility (Fig. 1b). This system is derived from the bacterial defense system where the CRISPR–Cas system recognizes foreign DNA and the nuclease activity of Cas9, which is guided to a particular sequence by sgRNAs, cleaves the invading DNA. However, in order to modulate gene expression without altering the DNA sequence, the endonuclease activity of Cas9 is inactivated and instead linked to transcriptional or epigenetic modulators (CRISPR–dCas9) (Sander and Joung 2014).

To summarize, all three systems have in common a programmable DNA-binding platform, designed to recognize a specific genomic DNA sequence. Subsequently, for epigenetic editing, an epigenetic modulator (or a catalytic domain thereof) is recruited to the locus of interest by tethering the effector domain to the DNA-binding platform,

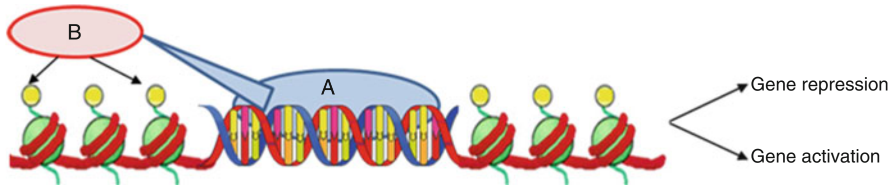


Fig. 2 Epigenetic editing is used to actively rewrite epigenetic signatures at a genomic location of interest. The molecular tools used for this purpose consist of (**a**) a DNA-binding platform to recognize the target sequence (see Fig. 1) and (**b**) an epigenetic modulator (or a catalytic domain thereof) which exerts its activating or repressive function by rewriting the epigenetic signature at a desired location

either directly (to ZF proteins or TALEs) or to the catalytically inactive Cas9 protein (which is recruited by sgRNAs). Upon delivery into target cells, the DNA-binding platform finds its DNA sequence, so the epigenetic modulator can expose its enzymatic activity at the desired genomic site (see Figs. 1 and 2). The epigenetic editing approach faced much disbelief in its early days, as the epigenetic marks were not generally considered to be instructive for gene expression, and if so, it was expected that their effect would be overruled by the native chromatin environment on a longer term. Moreover, the generally accepted inaccessibility of heterochromatic genes was thought to hamper successful editing of silenced genes. Pioneering studies by us and others and the introduction of straightforward DNA-targeting approaches set the stage for the recent boom in epigenetic editing (Jurkowski et al. 2015).

The by far most studied epigenetic mark is DNA methylation. It predominately occurs on cytosine followed by a guanine (CpG) sites; however, also non-CpG methylation has been detected in stem cells (Lister et al. 2009) and in the brain (Lister et al. 2013). In promoter regions, CpG dinucleotides often cluster in so-called CpG islands (CGIs), and more than half of the human gene promoters contain a CGI (Ehrlich et al. 1982; Saxonov et al. 2006). These CpG-rich promoters are usually unmethylated, with a few exceptions, including tissue-specific methylation during development (Bird et al. 1985; Song et al. 2005). Gene promoters found with high levels of DNA methylation are generally transcriptionally inactive (Boyes and Bird 1992; Siegfried et al. 1999; Jones and Takai 2001). Epigenetic editors will be very helpful to investigate whether methylation precedes gene inactivation or whether it is rather a consequence of inactivation, since (de)methylation can now be induced at will at specific genomic sites. This will also shed light on the order of events during the process of DNA methylation at, e.g., promoters, gene bodies, or enhancers.

2 Targeted DNA Methylation

Pioneering work in the field of targeted DNA methylation has been performed by Xu and Bestor who were the first to use a fusion protein consisting of an engineered ZF protein and a DNA methyltransferase to target DNA methylation to a predefined

DNA sequence (Xu and Bestor 1997). Several other studies of targeted DNA methylation using ZF proteins fused to human or bacterial DNA methyltransferases have been published, showing that the induction of DNA methylation indeed results in transcriptional repression (Smith and Ford 2007; Li et al. 2006, 2007; Minczuk et al. 2006; Smith et al. 2008; Carvin et al. 2003; McNamara et al. 2002; van der Gun et al. 2010). However, these early studies have only been performed on exogenous or nonmammalian sites. Endogenous gene repression by targeted DNA methylation was shown for the first time, in 2012 for the human gene promoters of *VEGF-A* (Siddique et al. 2012) *SOX2* and *MASPIN* (Rivenbark et al. 2012). Both studies used designed ZF proteins, engineered to bind a stretch of 18 bps within the promoter of the intended target gene. The ZF proteins were fused to the catalytic domain of the murine and human DNA methyltransferase 3A (DNMT3A) and a fusion of murine Dnmt3a or human DNMT3L, respectively.

The mouse Dnmt3a fusion resulted in a mean yield of 14.4 % DNA methylation over all CpG sites at the interrogated region of the *VEGF-A* promoter leading to a downregulation of mRNA expression by 36 % (Siddique et al. 2012). For certain CpGs, the induced methylation even reached efficiencies of 100 %. This average effect was further improved – up to a mean of 28.6 % DNA methylation and 56 % *VEGF-A* mRNA downregulation – when the effector domain consisted of the C-terminal domain of Dnmt3a fused to DNMT3L. This finding further proved that DNMT3L stimulates *de novo* methylation through Dnmt3a, as DNMT3L has no catalytic activity itself (Gowher et al. 2005).

Targeted DNA methylation of the tumor suppressor gene *MASPIN* using the catalytic domain of the human DNMT3A (598–908 amino acids) increased DNA methylation up to 60 % at single CpGs within the *MASPIN* promoter (Rivenbark et al. 2012). Increase of DNA methylation was detectable up to 500 bps downstream of the ZF target site and translated into a 50 % downregulation of mRNA and protein expression compared to an empty vector control. As expected, the downregulation of the tumor suppressor gene resulted in an increased proliferation rate and a more aggressive phenotype of breast cancer cells in vitro. In addition, the transcription factor *SOX2* was targeted using an inducible ZF–DNMT3A fusion (Rivenbark et al. 2012). Cell lines were stably transduced with the ZF–DNMT3A fusion, and as a control the same ZF protein fused to the transient repressor SKD (Kruppel-associated box domain) was used. This system is induced upon addition of doxycycline (Dox) to the culture medium, which causes expression of the fusion proteins. In turn, discontinuation of the Dox treatment led to depletion of ZF–DNMT3A and ZF–SKD expression, respectively. The expression of the *SOX2*-targeted ZF–DNMT3A construct translated into a 60–80 % downregulation of mRNA and protein expression, respectively. In a subsequent cell proliferation assay, the ZF–SKD construct was included as a control and both, ZF–SKD and ZF–DNMT3A, were initially able to decrease cell growth. However, when Dox was removed from the culture media 48 h after induction, only the ZF–DNMT3A fusion was able to attenuate cell proliferation over the time course of the experiment, suggesting stable gene repression mediated by DNA methylation, although DNA methylation at the *SOX2* promoter was not directly shown (Rivenbark et al. 2012). In a follow-up study, the

same lab showed that the silencing of *SOX2* expression was indeed mediated by targeted DNA methylation (Stolzenburg et al. 2015). Furthermore, depletion of Dox and subsequent discontinuation of the expression of the ZF–SKD and ZF–DNMT3A fusions led to the re-expression of *SOX2* only in ZF–SKD-transduced cells but not in cells that previously expressed the DNMT3A construct. Therefore, in this context, DNMT3A was a more stable mediator of expression than the SKD. Interestingly, *SOX2* mRNA and protein repression was stronger 8 days after Dox removal than the initial downregulation, suggesting that DNA methylation is reinforced by cellular mechanisms during subsequent cell divisions.

Using the mouse Dnmt3a fused to a ZF protein targeting the cell adhesion molecule *EpCAM* (van der Gun et al. 2013), an increase in DNA methylation at the endogenous *EpCAM* promoter by 20–25 % after transient transfection of an ovarian cancer cell line was observed (Nunna et al. 2014). At specific CpGs an increase of DNA methylation of more than 80 % was detected. In addition, two cell lines were generated that stably express ZF–Dnmt3a. Importantly, both cell lines showed an increase of DNA methylation at the *EpCAM* promoter of more than 40 %. The induction of promoter DNA methylation decreased the expression of *EpCAM* mRNA (60–70 %) and protein (50 %) in the examined cell line, and furthermore, the reduction of EpCAM expression translated into a decrease of the proliferative character of the examined ovarian cancer cell line.

Researchers have also employed TALEs for targeted DNA methylation studies, fused to either DNMT3A (Li et al. 2015) or DNMT3A–DNMT3L (Bernstein et al. 2015), targeting the promoters of *CRMP4* and *CDKN2A*, respectively. Both studies successfully showed induced DNA methylation at their respective TALE target sites. In both studies, the DNA methylation was associated with target gene repression and resulted in the intended physiological downstream effects. Although for the *CRMP4* promoter, DNA methylation only increased to about 5.5 % upstream to 6.4 % downstream of the TALE target site (numbers represent mean values of the interrogated region, with peaks of max 9–12 % at individual CpGs). The induced methylation, however, was sufficient to virtually completely knock down mRNA and protein expression in a nonmetastatic prostate cancer cell line (Li et al. 2015). This targeted DNA methylation was then shown to spread over 300 bps up- and downstream of the TALE–DNMT3A binding site. Importantly, the downregulation of *CRMP4*, a metastasis suppressor gene in a nonmetastatic cell line, led to the formation of metastasis in vivo. Furthermore, the crucial impact of DNA methylation at the *CRMP4* promoter for prostate cancer patients' survival was shown by DNA methylation analysis of prostate cancer specimen. The analysis revealed that 64 % of *CRMP4* methylation positive samples were indeed confirmed as metastatic.

Bernstein et al. (2015) engineered a TALE–Dnmt3a–DNMT3L (TALE–DNMT) construct to target the *CDKN2A* locus in HeLa cells, primary human fibroblasts, and coronary artery smooth muscle cells. The *CDKN2A* locus encodes the cyclin-dependent kinase inhibitor p16, a tumor suppressor, which is regulated by DNA methylation. The induced DNA methylation across the *CDKN2A* CpG island varied between 10 % (human fibroblasts) and 13.8 % (HeLa cells) after lentiviral transduction and 17 % in HeLa cells after sorting for successfully transfected cells. At

individual CpGs, the methylation levels increased even up to 66 % in the sorted population after transfection and 30–50 % after lentiviral transduction. The TALE–DNMT3A-mediated DNA methylation was associated with a 50 % decrease in *p16* mRNA expression in human fibroblasts accompanied by an increase in cell cycle progression. Recently, another group demonstrated ZF-induced methylation of the *CDKN2A* locus to promote migration and invasion of cancer cells (Cui et al. 2015).

These publications demonstrate that induction of endogenous DNA methylation at will by epigenetic editing tools at a specific target is not only possible but also effective, as treatment results in the intended physiological downstream effects. The induced DNA methylation needed to downregulate gene expression varied highly between the studies, and in one case as little as 10 % was sufficient to achieve target gene suppression. This strengthens the notion that a single CpG can be crucial for gene regulation at a given locus (Pogribny et al. 2000).

3 Stability of the Induced DNA Methylation Changes

It is well documented that promoter DNA methylation plays an important role in permanent gene silencing and that established DNA methylation is maintained during cell divisions to achieve stable gene repression (Riggs 1975; Holliday and Pugh 1975; Lister et al. 2009; Chen et al. 2007; Stein et al. 1982). However, nowadays it is presumed that DNA methylation by itself might not be enough to maintain stable gene repression in any given context. It is much more agreed that DNA methylation and a myriad of additional epigenetic mechanisms, such as histone modifications, nucleosome positioning, ncRNAs, and others, work together to create a stable context-dependent gene repression pattern (Raynal et al. 2012). Epigenetic editing provides unique tools to address sustainability of DNA methylation in different chromatin contexts.

The first study to address maintenance of written DNA methylation marks was performed by the Blancafort team. Toward this end, engineered ZF proteins targeting the *MASPIN* gene were fused to DNMT3A and retrovirally delivered into breast cancer cells (Rivenbark et al. 2012). The downregulation of the tumor suppressor *MASPIN* would lead to a more aggressive phenotype of the host cells. To prove this hypothesis, retrovirally transduced cells were seeded in soft agar for colony formation. After colonies were formed (several weeks later), single colonies were picked from the soft agar, disrupted, and cultured for subsequent sodium bisulfite sequencing to investigate the methylation state of *MASPIN*. The data revealed that even 50 days post-transduction DNA, methylation was maintained in the host cells (Rivenbark et al. 2012). Interestingly, knockdown of *UHRF1* (ubiquitin-like containing PHD and RING finger domains 1, a protein required for the maintenance of DNA methylation patterns (Bostick et al. 2007; Sharif et al. 2007)) led to re-expression of *MASPIN* in these cells (Rivenbark et al. 2012).

The longevity of the induced DNA methylation was further tested in vivo for the *SOX2* promoter in a xenograft mouse model (Stolzenburg et al. 2015). To do so, the advantages of the Dox-inducible system were exploited: the Dox-inducible system

allows controlled expression of the ZF–DNMT3A fusion by administration of a Dox-containing diet, whereas the switch to a Dox-free diet leads to the discontinuation of the ZF–DNMT3A expression. The results showed a strong tumor growth inhibition in the cells that expressed the ZF–DNMT3A fusion. This was associated with DNA methylation at the *SOX2* promoter together with a decrease in *SOX2* expression. Although DNA methylation at the *SOX2* promoter was largely sustained for 53 days post-Dox removal, after the removal of ZF–DNMT3A expression, tumor growth inhibition was only maintained for 10 days. Interestingly, examination of ZF–DNMT3A and *SOX2* expression at day 10 after Dox removal revealed a maintained repression of *SOX2* and no detection of ZF–DNMT3A, implying – once being induced – a long-term effect of DNA methylation on *SOX2* repression.

However, to truly verify the long-term effect of written DNA methylation signatures, DNA methylation and target gene expression should be validated at later time points. This notion is underpinned by a recent publication showing that ZF-targeted DNA methylation at the *VEGF-A* promoter by means of transient adenoviral transfer was not stably maintained (Kungulovski et al. 2015). The authors examined, after targeting a ZF–DNMT3A fusion to the *VEGF-A* promoter, the longevity of the induced DNA methylation mark at the *VEGF-A* promoter and *VEGF-A* expression over a time course of 15 days. In contrast to Stolzenburg et al., the loss of experimental ZF–DNMT3A expression was associated with a loss of DNA methylation at the target site and target gene re-expression (Kungulovski et al. 2015). Interestingly, the authors also looked into secondary effects of the induced DNA methylation on histone modifications. After induction of the targeted DNA methylation at the *VEGF-A* promoter, the authors examined whether the methylation mark at the DNA level is reinforced by a change in the silencing mark H3K9me3. However, no changes in H3K9me3 were detected using ChIP–qPCR (Kungulovski et al. 2015). As epigenetic editing is uniquely suited to address the parameters allowing or preventing maintenance of DNA methylation, ongoing research efforts are expected to yield important insights in this respect.

4 Targeted DNA Demethylation

Epigenome-wide association studies result in increasing lists of aberrantly hypermethylated loci associated with various clinical phenotypes. Mimicking these methylation profiles by epigenetic editing will provide valuable insights into the biological function of these modifications. More importantly, the actual removal of such epimutations will open new therapeutic avenues. Indeed, in the clinical setting, inhibitors of DNA methyltransferases are used to prevent hypermethylation of tumor suppressor genes. Unfortunately, such conventional epigenetic drugs will affect methylation patterns genome-wide. In contrast, epigenetic editing approaches might exploit the reversibility of epigenetic marks in a gene-targeted manner and in this way avoid potentially dangerous side effects.

Before the identification of active DNA methylation-modifying enzymes, DNA repair mechanisms were exploited for their role in reducing local DNA methylation

profiles. Indeed, Gregory et al. reported that the targeting of thymidine–DNA glycosylase (TDG) by fusion to engineered ZF proteins did result in lowering of DNA methylation, allowing improved induction of the target gene *Nos2* (Gregory et al. 2013). The identification of ten–eleven translocation (TET) enzymes and their role in modifying methylated cytosines allowed epigenetic editing approaches to actively reduce hypermethylation states of target genes without introducing temporary changes to the DNA. Indeed, we were the first group to report on the potency of targeting the TET domains to induce active DNA demethylation (Rots and Petersen-Mahrt 2013; Chen et al. 2014). Targeting either of the three TET members to the hypermethylated *ICAM* gene demonstrated that both TET1 and TET2 are effective reducers of DNA methylations. Although DNA demethylation in this experimental setting was relatively low (minus ca 5 approx.), gene expression was increased two-fold. Obviously, when compared to targeting of VP64, a strong viral transcription activator, this gene expression modulation was modest. However, such mild increases might be physiologically more relevant, and these findings do generate opportunities to realize therapeutically relevant localized DNA demethylation. Indeed, we confirmed the robustness of the TET2-targeting approach in inducing DNA demethylation for four other genes (*EPB41L3*, *C13ORF18*, *CCNA1*, and *TFPI2*; all putative hypermethylation biomarkers for cervical cancer) (Huisman et al. 2015a; Huisman et al. 2015b). Although the observation that modest local demethylation is less effective in gene re-expression than targeting a transcriptional activator is understandable, the large size of TET domains is also partially responsible for this effect. Interestingly, when cells were co-treated with the epigenetic drug trichostatin A (TSA), which is a histone deacetylase inhibitor that might increase gene accessibility, induced expression of silenced target genes was detectable (Huisman et al. 2015a).

Despite the large size of the TET domain, also in fusion with the relatively larger TALE domains, TET1 was able to induce targeted DNA demethylation (Maeder et al. 2013a; Li et al. 2015). Maeder et al. were the first to show targeted demethylation using TALEs fused to TET1. In total, they engineered 25 TET1-containing TALEs targeting *KLF4*, *HBB*, and *RHOXF2* (Maeder et al. 2013a). Comparison of the TALE constructs – either fused to the full-length TET1 or the constructs containing only the TET1 catalytic domain (TET1c) – showed that the catalytic domain had a stronger effect on demethylating its target genes (up to 30 % at *KLF4* and even 84 % at *HBB*) than the full-length TET1 (Maeder et al. 2013a). The most effective demethylation was observed within 30 bps up- and downstream but also up to 200 bps away from the target sequence. This is in accordance with a study published by Li et al., who examined two regions located 4 bps upstream and 95 bps downstream of the TALE–TET1c target site for their methylation status. Both regions showed demethylation with the more distant region being more efficiently demethylated. In this study, DNA demethylation was associated with a re-expression of target gene mRNA expression, followed by an induction of protein expression (Li et al. 2015). As expected, the demethylation and re-expression of *CRMP4* (a metastasis suppressor gene) showed a decrease in migration and invasion in otherwise metastatic cell lines. Furthermore, the re-expression of *CRMP4* after active DNA

demethylation abolished the metastatic character of these cells even in an in vivo mouse model of prostate cancer (Li et al. 2015). However, demethylation of the target gene by TALE–TET1c did not always result in induced gene expression. Only four out of ten demethylating TALE–TET1c constructs targeting *HBB* indeed increased the expression of HBB mRNA. Similarly, in the case of *RHOXF2*, two out of five demethylating TALEs–TET1c induced mRNA expression (Maeder et al. 2013a). The authors suggested that the artificial demethylation at the target gene was not stably transmitted, and therefore, demethylated CpGs became remethylated, as TALE–TET1c coding constructs became cleared from the transfected cells.

A very interesting aspect of the targeted DNA demethylation, namely, its effect on histone modifications, was examined by Li et al. for both regions that showed a decrease in TALE–TET1c-mediated DNA methylation (Li et al. 2015). While the region directly upstream to the TALE–TET1c target site showed a reduction of repressive histone modifications (H3K9me3, H3K27me3, H3K79me3), this was not seen at the region 95 bps downstream of the target site, although this region showed stronger demethylation (Li et al. 2015).

5 Concluding Remarks and Future Perspectives

Epigenetics has been receiving a lot of attention in this post-genomic era: many abnormalities in the epigenetic landscape have been identified in numerous diseases, and so-called epigenetic drugs, including inhibitors of DNA methyltransferase, have entered the clinical arena. Epigenetic editing – to mimic or reverse such epimutations – is currently gaining widespread acceptance, and many research groups join the field. Initially, the technology had to overcome some hurdles: genome specificity was not likely to be achieved, accessibility of silenced genes was thought to be impossible, and the instructive nature of epigenetic marks with respect to controlling gene expression was highly questioned. As reviewed here, these assumptions have been proven untrue. To increase the specificity of targeting, a considerable progress has been made in the field of enzyme engineering, where split enzymes allow activity to take place only when two split parts are brought close together via their fusion to two closely binding DNA-targeting modules (Kiss and Weinhold 2008). Genome-wide specificity can also be achieved using CRISPR–dCas9 technology (Hilton et al. 2015), and the progress in sgRNA design is expected to rapidly improve our understanding of the off-target effects. Taken together, this allows the prediction and prevention of unwanted side effects due to unintended endogenous binding.

Also – against common belief – heterochromatin is not hampering accessibility per se, as re-expression of silenced genes has now been shown for many heterochromatic genes (tumor suppressors) by ATFs (Beltran et al. 2007; Lara et al. 2012; van der Wijst et al. 2015; Falahi et al. 2013; Huisman et al. 2013) and epigenetic editors (e.g. Chen et al. 2014; Huisman et al. 2015a, b). Although the transcriptional activators/repressors in ATFs are relatively small in size, also larger constructs can gain gene access even though size is likely to affect effectiveness for heterochromatic genes.

On top of that, the heterochromatin landscape, which is unique for each gene and cell, needs further investigation in order to completely understand the mechanism of action of different epigenetic editors.

With respect to the cause versus consequence relationship of epigenetic marks and gene expression, strong indications that epigenetics is instructive in gene expression regulation have been obtained by targeting effector domains to artificial loci (e.g., plasmids and integrated sites) as reviewed by us in 2012 (de Groote et al. 2012). At that time, only two examples were published, which confirmed that editing of epigenetic marks at a predetermined endogenous site was effective in modulating gene expression (Snowden et al. 2002; Rivenbark et al. 2012). These days, epigenetic/epigenome editing has been declared a method to watch (Rusk 2014), and a rapid increase in publications confirms efficiency of the approach (Ledford 2015). An important open question concerns the chromatin microenvironment conditions allowing sustained re-expression, but the technology of epigenetic editing is uniquely qualified to address this question (Cano-Rodriguez et al. 2016). As also the CRISPR–dCas9 platform is currently exploited for targeted (de)methylation (Choudhury et al. 2016; McDonald et al. 2016; Vojta et al. 2016; Xu et al. 2016), we expect epigenetic editing tools to soon reprogram the genome in a sustained manner, which will provide a clinically relevant hit-and-run approach to cure currently incurable diseases, including imprinting (Bashtrykov et al. 2015) or behavioral disorders (Dekker et al. 2014).

Acknowledgments We would like to acknowledge the EU funding for D.G. (H2020-MSCA-ITN-2014-ETN 642691 EpiPredict). M.G.R. serves as vice-chair of H2020-COST CM1406, and her team is partially funded by NWO-Vidi-91786373 and EU-FP7-SNN-4D22C-T2007.

References

Bashtrykov P, Kungulovski G, Jeltsch A. Correction of aberrant imprinting by allele specific epigenome editing. Clin Pharmacol Ther. 2015;99(5):482–4.

Beltran A, Parikh S, Liu Y, Cuevas BD, Johnson GL, Futscher BW, et al. Re-activation of a dormant tumor suppressor gene maspin by designed transcription factors. Oncogene. 2007;26(19):2791–8.

Bernstein DL, Le Lay JE, Ruano EG, Kaestner KH. TALE-mediated epigenetic suppression of CDKN2A increases replication in human fibroblasts. J Clin Invest. 2015;125(5):1998–2006.

Bird A, Taggart M, Frommer M, Miller OJ, Macleod D. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell. 1985;40(1):91–9.

Boch J, Bonas U. Xanthomonas AvrBs3 family-type III effectors: discovery and function. Annu Rev Phytopathol. 2010;48:419–36.

Bostick M, Kim JK, Esteve PO, Clark A, Pradhan S, Jacobsen SE. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science. 2007;317(5845):1760–4.

Boyes J, Bird A. Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. EMBO J. 1992;11(1):327–33.

Briggs AW, Rios X, Chari R, Yang L, Zhang F, Mali P, et al. Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. Nucleic Acids Res. 2012;40(15):e117.

Cano-Rodriguez D, Gjaltema RA, Jilderda LJ, Jellema P, Dokter-Fokkens J, Ruiters MH, Rots MG. Writing of H3K4Me3 overcomes epigenetic silencing in a sustained but context-dependent manner. Nat Commun. 2016;7:12284. (doi:10.1038/ncomms12284. PubMed PMID: 27506838).

Carvin CD, Parr RD, Kladde MP. Site-selective in vivo targeting of cytosine-5 DNA methylation by zinc-finger proteins. Nucleic Acids Res. 2003;31(22):6493–501.

Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, Schmidt C, et al. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res. 2011;39(12):e82.

Chen H, Kazemier HG, de Groote ML, Ruiters MH, Xu GL, Rots MG. Induced DNA demethylation by targeting Ten-Eleven Translocation 2 to the human ICAM-1 promoter. Nucleic Acids Res. 2014;42(3):1563–74.

Chen T, Hevi S, Gay F, Tsujimoto N, He T, Zhang B, et al. Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. Nat Genet. 2007;39(3):391–6.

Choo Y, Sanchez-Garcia I, Klug A. In vivo repression by a site-specific DNA-binding protein designed against an oncogenic sequence. Nature. 1994;372(6507):642–5.

Choudhury SR, Cui Y, Lubecka K, Stefanska B, Irudayaraj J. CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1 promoter. Oncotarget. 2016. (Epub ahead of print) (doi:10.18632/oncotarget.10234. PubMed PMID: 27356740).

Cui C, Gan Y, Gu L, Wilson J, Liu Z, Zhang B, et al. P16-specific DNA methylation by engineered zinc finger methyltransferase inactivates gene transcription and promotes cancer metastasis. Genome Biol. 2015;16(1):252.

de Groote ML, Verschure PJ, Rots MG. Epigenetic Editing: targeted rewriting of epigenetic marks to modulate expression of selected target genes. Nucleic Acids Res. 2012;40(21):10596–613.

Dekker AD, De Deyn PP, Rots MG. Epigenetics: the neglected key to minimize learning and memory deficits in Down syndrome. Neurosci Biobehav Rev. 2014;45:72–84.

Ehrlich M, Gama-Sosa MA, Huang LH, Midgett RM, Kuo KC, McCune RA, et al. Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. Nucleic Acids Res. 1982;10(8):2709–21.

Falahi F, Huisman C, Kazemier HG, van der Vlies P, Kok K, Hospers GA, et al. Towards sustained silencing of HER2/neu in cancer by epigenetic editing. Mol Cancer Res. 2013;11(9):1029–39.

Falahi F, Sgro A, Blancafort P. Epigenome engineering in cancer: fairytale or a realistic path to the clinic? Front Oncol. 2015;5:22.

Gaj T, Gersbach CA, Barbas 3rd CF. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol. 2013;31(7):397–405.

Gowher H, Liebert K, Hermann A, Xu G, Jeltsch A. Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. J Biol Chem. 2005;280(14):13341–8.

Gregory DJ, Zhang Y, Kobzik L, Fedulov AV. Specific transcriptional enhancement of inducible nitric oxide synthase by targeted promoter demethylation. Epigenetics. 2013;8(11):1205–12.

Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE, et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat Biotechnol. 2015;33(5):510–7.

Holliday R, Pugh JE. DNA modification mechanisms and gene activity during development. Science. 1975;187(4173):226–32.

Huisman C, van der Wijst MG, Falahi F, Overkamp J, Karsten G, Terpstra MM, et al. Prolonged re-expression of the hypermethylated gene EPB41L3 using artificial transcription factors and epigenetic drugs. Epigenetics. 2015a;10(5):384–96.

Huisman C, van der Wijst MG, Schokker M, Blancafort P, Terpstra MM, Kok K, et al. Re-expression of selected epigenetically silenced candidate tumor suppressor genes in cervical cancer by TET2-directed demethylation. Mol Ther. 2015b;24(3):536–47.

Huisman C, Wisman GB, Kazemier HG, van Vugt MA, van der Zee AG, Schuuring E, et al. Functional validation of putative tumor suppressor gene C13ORF18 in cervical cancer by Artificial Transcription Factors. Mol Oncol. 2013;7(3):669–79.

Jones PA, Takai D. The role of DNA methylation in mammalian epigenetics. Science. 2001;293(5532):1068–70.

Jurkowski TP, Ravichandran M, Stepper P. Synthetic epigenetics-towards intelligent control of epigenetic states and cell identity. Clin Epigenetics. 2015;7(1):18.

Kim CA, Berg JM. A 2.2 A resolution crystal structure of a designed zinc finger protein bound to DNA. Nat Struct Biol. 1996;3(11):940–5.

Kiss A, Weinhold E. Functional reassembly of split enzymes on-site: a novel approach for highly sequence-specific targeted DNA methylation. Chembiochem. 2008;9(3):351–3.

Kungulovski G, Jeltsch A. Epigenome editing: state of the art, concepts, and perspectives. Trends Genet. 2015;32(2):101–13.

Kungulovski G, Nunna S, Thomas M, Zanger UM, Reinhardt R, Jeltsch A. Targeted epigenome editing of an endogenous locus with chromatin modifiers is not stably maintained. Epigenetics Chromatin. 2015;8:12.

Lara H, Wang Y, Beltran AS, Juarez-Moreno K, Yuan X, Kato S, et al. Targeting serous epithelial ovarian cancer with designer zinc finger transcription factors. J Biol Chem. 2012;287(35): 29873–86.

Ledford H. Targeted gene editing enters clinic. Nature. 2011;471(7336):16.

Ledford H. Epigenetics: the genome unwrapped. Nature. 2015;528(7580):S12–3.

Li F, Papworth M, Minczuk M, Rohde C, Zhang Y, Ragozin S, et al. Chimeric DNA methyltransferases target DNA methylation to specific DNA sequences and repress expression of target genes. Nucleic Acids Res. 2007;35(1):100–12.

Li H, Rauch T, Chen ZX, Szabo PE, Riggs AD, Pfeifer GP. The histone methyltransferase SETDB1 and the DNA methyltransferase DNMT3A interact directly and localize to promoters silenced in cancer cells. J Biol Chem. 2006;281(28):19489–500.

Li K, Pang J, Cheng H, Liu WP, Di JM, Xiao HJ, et al. Manipulation of prostate cancer metastasis by locus-specific modification of the CRMP4 promoter region using chimeric TALE DNA methyltransferase and demethylase. Oncotarget. 2015;6(12):10030–44.

Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, et al. Global epigenomic reconfiguration during mammalian brain development. Science. 2013;341(6146):1237905.

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009; 462(7271):315–22.

Maeder ML, Angstman JF, Richardson ME, Linder SJ, Cascio VM, Tsai SQ, et al. Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. Nat Biotechnol. 2013a;31(12):1137–42.

Maeder ML, Linder SJ, Reyon D, Angstman JF, Fu Y, Sander JD, et al. Robust, synergistic regulation of human gene expression using TALE activators. Nat Methods. 2013b;10(3):243–5.

McDonald JI, Celik H, Rois LE, Fishberger G, Fowler T, Rees R, et al. Reprogrammable CRISPR/Cas9-based system for inducing site-specific DNA methylation. Biol Open. 2016;5(6):866–74. (doi:10.1242/bio.019067. PubMed PMID: 27170255; PubMed Central PMCID: PMC4920199).

McNamara AR, Hurd PJ, Smith AE, Ford KG. Characterisation of site-biased DNA methyltransferases: specificity, affinity and subsite relationships. Nucleic Acids Res. 2002;30(17): 3818–30.

Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, et al. A TALE nuclease architecture for efficient genome editing. Nat Biotechnol. 2011;29(2):143–8.

Minczuk M, Papworth MA, Kolasinska P, Murphy MP, Klug A. Sequence-specific modification of mitochondrial DNA using a chimeric zinc finger methylase. Proc Natl Acad Sci U S A. 2006; 103(52):19689–94.

Mussolino C, Morbitzer R, Lutge F, Dannemann N, Lahaye T, Cathomen T. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic Acids Res. 2011;39(21):9283–93.

Nunna S, Reinhardt R, Ragozin S, Jeltsch A. Targeted methylation of the epithelial cell adhesion molecule (EpCAM) promoter to silence its expression in ovarian cancer cells. PLoS One. 2014;9(1):e87703.

Pogribny IP, Pogribna M, Christman JK, James SJ. Single-site methylation within the p53 promoter region reduces gene expression in a reporter gene construct: possible in vivo relevance during tumorigenesis. Cancer Res. 2000;60(3):588–94.

Raynal NJ, Si J, Taby RF, Gharibyan V, Ahmed S, Jelinek J, et al. DNA methylation does not stably lock gene expression but instead serves as a molecular mark for gene silencing memory. Cancer Res. 2012;72(5):1170–81.
Reyon D, Tsai SQ, Khayter C, Foden JA, Sander JD, Joung JK. FLASH assembly of TALENs for high-throughput genome editing. Nat Biotechnol. 2012;30(5):460–5.
Riggs AD. X inactivation, differentiation, and DNA methylation. Cytogenet Cell Genet. 1975;14(1):9–25.
Rivenbark AG, Stolzenburg S, Beltran AS, Yuan X, Rots MG, Strahl BD, et al. Epigenetic reprogramming of cancer cells via targeted DNA methylation. Epigenetics. 2012;7(4):350–60.
Rots MG, Petersen-Mahrt SK. The 2012 IMB Conference: DNA demethylation, repair and beyond. Institute of Molecular Biology, Mainz, Germany, 18-21 October 2012. Epigenomics. 2013; 5(1):25–8.
Rusk N. CRISPRs and epigenome editing. Nat Methods. 2014;11(1):28.
Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol. 2014;32(4):347–55.
Saxonov S, Berg P, Brutlag DL. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proc Natl Acad Sci U S A. 2006;103(5): 1412–7.
Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, Endo TA, et al. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature. 2007; 450(7171):908–12.
Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. Carcinogenesis. 2010;31(1):27–36.
Siddique AN, Nunna S, Rajavelu A, Zhang Y, Jurkowska RZ, Reinhardt R, et al. Targeted Methylation and Gene Silencing of VEGF-A in Human Cells by Using a Designed Dnmt3a-Dnmt3L Single-Chain Fusion Protein with Increased DNA Methylation Activity. J Mol Biol. 2012;425(3):479–91.
Siegfried Z, Eden S, Mendelsohn M, Feng X, Tsuberi BZ, Cedar H. DNA methylation represses transcription in vivo. Nat Genet. 1999;22(2):203–6.
Smith AE, Ford KG. Specific targeting of cytosine methylation to DNA sequences in vivo. Nucleic Acids Res. 2007;35(3):740–54.
Smith AE, Hurd PJ, Bannister AJ, Kouzarides T, Ford KG. Heritable gene repression through the action of a directed DNA methyltransferase at a chromosomal locus. J Biol Chem. 2008; 283(15):9878–85.
Snowden AW, Gregory PD, Case CC, Pabo CO. Gene-specific targeting of H3K9 methylation is sufficient for initiating repression in vivo. Curr Biol. 2002;12(24):2159–66.
Song F, Smith JF, Kimura MT, Morrow AD, Matsuyama T, Nagase H, et al. Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. Proc Natl Acad Sci U S A. 2005;102(9):3336–41.
Stein R, Razin A, Cedar H. In vitro methylation of the hamster adenine phosphoribosyltransferase gene inhibits its expression in mouse L cells. Proc Natl Acad Sci U S A. 1982;79(11): 3418–22.
Stolzenburg S, Beltran AS, Swift-Scanlan T, Rivenbark AG, Rashwan R, Blancafort P. Stable oncogenic silencing in vivo by programmable and targeted de novo DNA methylation in breast cancer. Oncogene. 2015;34(43):5427–35.
van der Gun BT, Huisman C, Stolzenburg S, Kazemier HG, Ruiters MH, Blancafort P, et al. Bidirectional modulation of endogenous EpCAM expression to unravel its function in ovarian cancer. Br J Cancer. 2013;108(4):881–6.
van der Gun BT, Maluszynska-Hoffman M, Kiss A, Arendzen AJ, Ruiters MH, McLaughlin PM, et al. Targeted DNA methylation by a DNA methyltransferase coupled to a triple helix forming oligonucleotide to down-regulate the epithelial cell adhesion molecule. Bioconjug Chem. 2010;21(7):1239–45. doi:10.1021/bc1000388.
van der Wijst MG, Huisman C, Mposhi A, Roelfes G, Rots MG. Targeting Nrf2 in healthy and malignant ovarian epithelial cells: protection versus promotion. Mol Oncol. 2015;9(7):1259–73.

Vandevenne M, Jacques DA, Artuz C, Nguyen CD, Kwan AH, Segal DJ, et al. New insights into DNA recognition by zinc fingers revealed by structural analysis of the oncoprotein ZNF217. J Biol Chem. 2013;288(15):10616–27.

Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. A census of human transcription factors: function, expression and evolution. Nat Rev Genet. 2009;10(4):252–63.

Vojta A, Dobrinić P, Tadić V, Bočkor L, Korać P, Julg B, et al. Repurposing the CRISPR-Cas9 system for targeted DNA methylation. Nucleic Acids Res. 2016;44(12):5615–28. (doi:10.1093/nar/gkw159. Epub 2016 Mar 11; PMID: 26969735; PubMed Central PMCID: PMC4937303).

Waddington CH. The epigenotype. 1942. Int J Epidemiol. 2012;41(1):10–3.

Xu GL, Bestor TH. Cytosine methylation targetted to pre-determined sequences. Nat Genet. 1997;17(4):376–8.

Xu X, Tao Y, Gao X, Zhang L, Li X, Zou W, et al. A CRISPR- based approach for targeted DNA demethylation. Cell Discov. 2016;3;2:16009. (doi:10.1038/celldisc.2016.9. eCollection 2016. PubMed PMID: 27462456; PubMed Central PMCID: PMC4853773).

Engineering and Directed Evolution of DNA Methyltransferases

Paola Laurino, Liat Rockah-Shmuel, and Dan S. Tawfik

Abstract

DNA methyltransferases (MTases) constitute an attractive target for protein engineering, thus opening the road to new ways of manipulating DNA in a unique and selective manner. Here, we review various aspects of MTase engineering, both methodological and conceptual, and also discuss future directions and challenges. Bacterial MTases that are part of restriction/modification (R/M) systems offer a convenient way for the selection of large gene libraries, both in vivo and in vitro. We review these selection methods, their strengths and weaknesses, and also the prospects for new selection approaches that will enable the directed evolution of mammalian DNA methyltransferases (Dnmts). We explore various properties of MTases that may be subject to engineering. These include engineering for higher stability and soluble expression (MTases, including bacterial ones, are prone to misfolding), engineering of the DNA target specificity, and engineering for the usage of S-adenosyl-L-methionine (AdoMet) analogs. Directed evolution of bacterial MTases also offers insights into how these enzymes readily evolve in nature, thus yielding MTases with a huge spectrum of DNA target specificities. Engineering for alternative cofactors, on the other hand, enables modification of DNA with various groups other than methyl and thus can be employed to map and redirect DNA epigenetic modifications.

P. Laurino • L. Rockah-Shmuel • D.S. Tawfik (✉)
Department of Biological Chemistry, Weizmann Institute of Science,
Rehovot 76100, Israel
e-mail: dan.tawfik@weizmann.ac.il

© Springer International Publishing Switzerland 2016

A. Jeltsch, R.Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*,
Advances in Experimental Medicine and Biology 945,
DOI 10.1007/978-3-319-43624-1_18

Abbreviations

MTase	DNA methyltransferase
Dnmt	Mammalian DNA methyltransferase
R/M	Restriction/modification
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IVC	In vitro compartmentalization
PCR	Polymerase chain reaction
MeDIP	Methylated DNA immunoprecipitation
CpG	5’-C-phosphate-G-3’
NMR	Nuclear magnetic resonance
ELISA	Enzyme-linked immunosorbent assay
DIG	Digoxigenin
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis

1 Introduction

DNA methyltransferases belong to a large and highly diverse group of enzymes that modify DNA. Given the potential utility of enzymes working on DNA in genetic and genomic engineering, in genome analysis, and in gene therapy, enzymes such as polymerases, ligases, and recombinases have been subjected to engineering, by rational design and/or by directed evolution – namely, by generation of gene libraries containing random mutations and their selection for mutants that improve the trait under selection. MTases are a critical part of nature’s DNA modification toolbox, both in prokaryotes, primarily as part of restriction/modification (R/M) systems, and in eukaryotes, primarily as epigenetic modifiers. As such, they constitute an attractive target for enzyme engineering.

Directed evolution involves the design and generation of gene libraries followed by a screen or a selection step. An attractive feature of DNA-modifying enzymes is the relative ease of selection, as cycles of selection can be applied, whereby the enzyme variants that encode the desired property modify their own encoding genes in a way that enables their isolation from non-modified genes (that in turn encode enzymes that do not possess the desired function). A prerequisite for such cycles is a linkage between the genotype (the gene encoding the enzyme) and the phenotype (the function exerted by the encoded enzyme). As discussed in Section 2, in the case of MTases, this linkage can be achieved either *in vivo* (in living cells) or in vitro (in artificial cells).

While this chapter provides a detailed overview of the selection strategies that are applicable to MTases, the design and construction of DNA libraries are not discussed in detail. There are numerous approaches for the generation of gene libraries. Most of these approaches are generic and can also be applied to MTases. We thus refer the reader to recent reviews, including ours (Goldsmith and Tawfik 2013; Rockah-Shmuel et al. 2014; Miyazaki 2003; Ulrich et al. 2012; Stemmer 1994;

Dalby 2011; Morley and Kazlauskas 2005; Jochens and Bornscheuer 2010; Hart and Waldo 2013). We elaborate, however, on the library design, aiming specifically at improved protein stability, which is a major limiting factor in engineering and evolving proteins in general and MTases in particular (Tokuriki and Tawfik 2009b). As manifested in very low yields of soluble and functional enzyme upon overexpression in *E. coli*, MTases, both bacterial and mammalian ones, seem to be prone to misfolding and aggregation. Section 3.1 addresses mutational strategies aimed at improving MTase stability.

In Sections 3.2 and 3.3, we explore various functional properties of MTases that may be subject to engineering. These include the engineering of new DNA target specificities and of new cofactor specificities which enable the usage of alternative cofactors that may serve either as methyl donors (similarly to the natural cofactor AdoMet) or that will lead to DNA modification with alternative alkyl groups.

Finally, Section 4 summarizes future prospects for MTase engineering, as well as the challenges that MTase engineers currently face. The latter include technical challenges, e.g., the development of selection methods for epigenetic Dnmts, specific for GpC, as well as conceptual ones, e.g., the challenge of obtaining full orthogonality, namely, the enzyme variants whose activity with the original DNA target, or cofactor (AdoMet), is effectively nil, while their activity with the new target, or cofactor, is comparable to the wild type with its original target and cofactor.

2 Methodology: Screening and Selection Methods

Directed evolution regards the selection of gene libraries, whereby the target enzyme is mutated at one or more positions. As mentioned above, selection demands a linkage between the gene encoding the enzyme and the phenotype that usually reflects enzyme's activity or other desired property. The obvious solution is compartmentalization, either in living cells or in artificial compartments.

2.1 Selection by In Vivo Plasmid Protection

Transformation of the gene library to bacteria, typically E. coli, results in individual bacteria carrying individual MTase genes. This strategy, by which bacterial MTases can be readily selected, was originally used for their functional cloning (Szomolanyi et al. 1980). Selection is based on plasmid protection against restriction endonuclease (Fig. 1a). Each round of evolution includes four steps: Diversification of the MTase gene by in vitro mutagenesis (*1*) and cloning of the library genes into an expression plasmid and transformation into *E. coli* (*2*) are two of these steps. This is followed by the growth of the transformed bacteria and expression of the MTase variant encoded by the plasmid. During this step, plasmids encoding the desired MTase are methylated at the respective restriction/modification (R/M) target sites (self-methylation) (*3*). The plasmid DNA is then extracted and digested with the

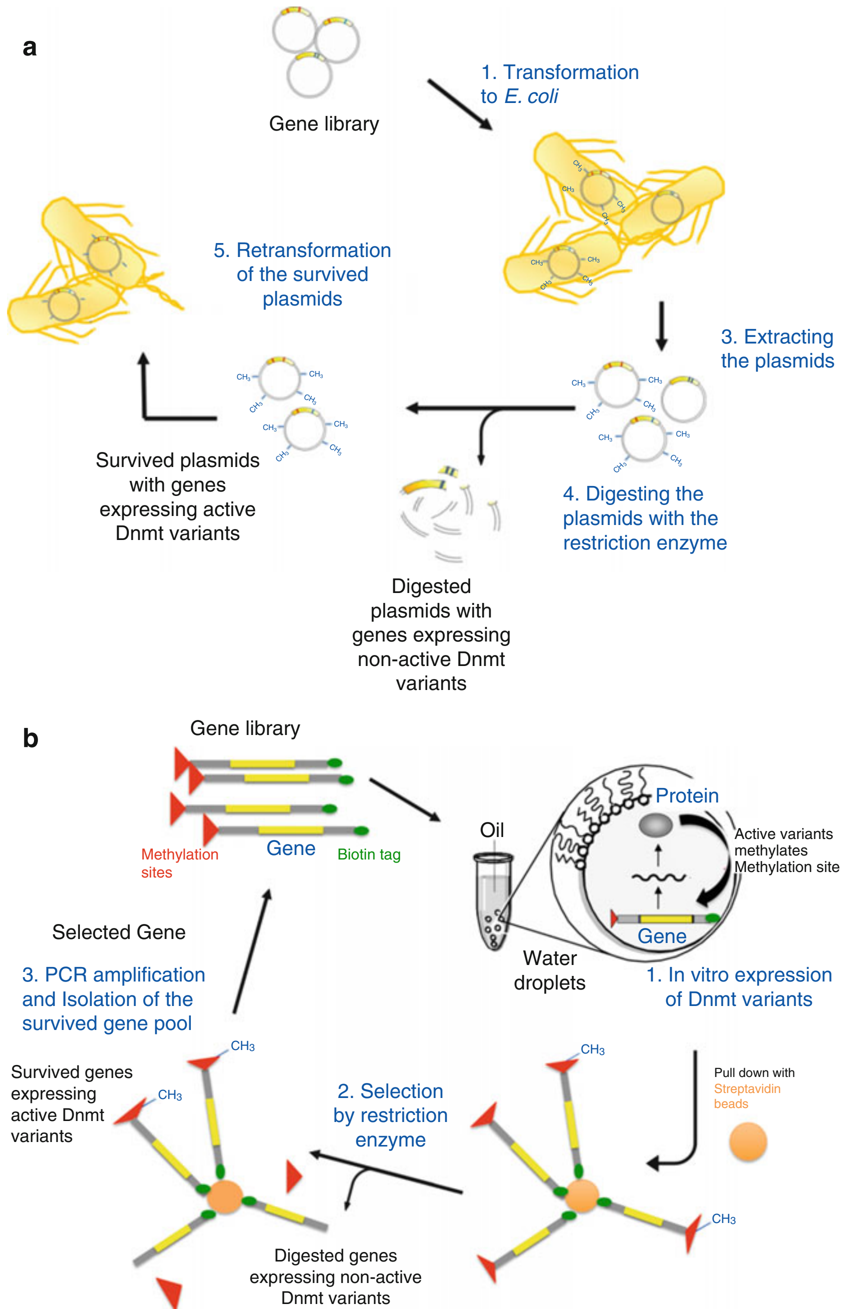


Fig. 1 Overview of the selection methods for MTases. (**a**) In vivo selection via plasmid protection. Extraction of the selected plasmids follows step 5; then the selected genes are sequenced to detect the improving mutations or are amplified by error-prone PCR, cloned back into the vector, and directed to another round of selection. (**b**) Selection by in vitro compartmentalization

cognate restriction enzyme that is blocked by the methylation within its recognition sequence (*4*). The digested plasmid mixture is transformed, and only methylated plasmids that were fully protected from digestion give viable colonies.

The plasmid protection strategy has clear advantages and has been widely used (Rockah-Shmuel and Tawfik 2012; Jurkowska et al. 2011; Samuelson et al. 2006; Timar et al. 2004; Gerasimaite et al. 2009; Chahar et al. 2010; Jeltsch et al. 1996; Friedrich et al. 1998). Foremost, it is very easy to implement – all that is needed, basically, is a plasmid containing the desired R/M target sites and the gene encoding for the MTase. Secondly, the progress of selection can be readily monitored, simply by running the digestion mixture on a gel and examining the fractions of intact versus digested plasmid. Thirdly, the selection pressure can be easily tuned by controlling the expression levels of the MTase, typically, by reducing the levels of the expression inducer (be it IPTG or any other inducer) (Rockah-Shmuel and Tawfik 2012).

However, there are also obvious caveats associated with the in vivo plasmid protection selection approach. Cells are complex entities and often evolve mechanisms to circumvent the applied selection pressure. For example, point mutations or recombination in the plasmid may result in the loss of the R/M sites and thereby in protection irrespective of the plasmid encoding an active MTase. Few tips for setting up successful selection systems are as follows: (*1*) ensure to use multiple R/M sites per plasmid, (*2*) have a minimal number of R/M sites within the MTase's ORF and the promoter, and (*3*) include at least one, ideally more, R/M site in the antibiotics resistance gene to prevent the takeover by "selfish DNA" (Rockah-Shmuel and Tawfik 2012). The phenomenon of takeover by "selfish DNA" is further intensified by undesirable effects that MTases often have in bacterial hosts. Specifically, some methylation patterns are toxic and require the use of mutated *E. coli* strains (McrBC^{-}, deficient in an endonuclease which cleaves DNA containing methylcytosine on one or both strands); e.g., expression of M.HaeIII requires the strain ER2267 or MC1061 (Raleigh and Wilson 1986). Additionally, although MTases do not exhibit high turnover numbers, even at comparatively low MTase expression levels, the bacterial genome undergoes a complete or nearly complete methylation (Rockah-Shmuel and Tawfik 2012). Thus, the higher the activity of the evolving variant, the more toxic it may become for the host bacteria. Indeed, the methylation of the host genome may affect the selection and specifically may exert counter-selection pressures on the evolving MTase, leading to depletion of the most active variants. Indeed, certain MTase selections were found to enrich for stop codons and frameshifts (Rockah-Shmuel and Tawfik 2012).

The scope of this in vivo selection strategy is also limited to the target site dictated by the restriction enzyme. The plasmid protection strategy cannot be used, for example, for the selection of mammalian Dnmts that target CpG sites. In principle, a specific restriction enzyme whose R/M site contains CG within its target site can be applied. However, since CpG sites are so frequent, on the selection plasmid, not to mention the host genome, the applied Dnmt needs to be hyperactive to methylate all sites including the applied site. Indeed, when expressed in *E. coli*, human Dnmts barely show plasmid protection (Jurkowska et al. 2011).

Another general drawback of the in vivo systems is the library size, which is limited by the transformation efficiency. Obtaining up to 10^6 transformants is routine in *E. coli*. Much larger libraries can be transformed, even at levels of 10^{10} transformants, but this demands considerable effort. In other host organisms, be they bacteria or eukaryotes such as yeast, the transformation efficiency constitutes an even more significant hurdle.

Finally, the scope of in vivo selections is also limited with respect to directed evolution toward the usage of synthetic AdoMet analogs. Due to its essentiality, the natural cofactor, AdoMet, is present at relatively high concentrations in the cells (0.1–0.2 mM in *E. coli* (Bennett et al. 2009)). AdoMet concentrations can be reduced by expressing the enzyme AdoMet hydrolase (LaMonte and Hughes 2006; Hughes et al. 1987), but AdoMet cannot be completely removed from the bacterial host. Additionally, AdoMet, and presumably adoMet analogs, do not readily enter cells (Lin et al. 2001). Low permeability means that cellular concentration of any synthetic AdoMet analog added to the culture media may be orders of magnitude lower compared to AdoMet, thus resulting in AdoMet utilizing mutants dominating the selection even if their overall methylation efficiency is low.

2.2 Selection by In Vitro Compartmentalization

An alternative to the in vivo selection is the cell-free environment of an in vitro compartmentalization (IVC), which allows the selection of larger libraries ($\geq 10^{10}$). IVC is a technique for directed evolution of enzymes based on compartmentalizing the transcription and translation of the enzyme-coding genes and the selection for enzymatic reactions in droplets of water-in-oil emulsions (Fig. 1b). The in vitro selection of MTases follows the principle of plasmid self-methylation. Droplets containing active MTase variants methylate their own encoding gene, while in other compartments genes encoding inactive enzymes remain unmodified. Consequently, genes encoding active variants are selectively enriched by applying the cognate restriction enzyme (Bernath et al. 2005; Cohen et al. 2004; Griffiths and Tawfik 2003; Bernath et al. 2004).

IVC cannot only handle larger libraries, but it also removes the above-discussed complications and interferences that relate to MTase expression and methylation activity in living cells. The IVC selection pressure can be tuned, not only by altering the number of methylation sites but also by adding noncoding DNAs that carry different target sites and compete with the methylation sites on the MTase-coding DNA. Similarly, artificial cofactors can be used by applying NEB's Pure™ cell-free translation system that only contains those *E. coli* proteins that are essential for transcription and translation.

On the downside, in comparison to the in vivo selection, IVC is a relatively complex system to handle: (1) the MTase gene must efficiently express and be active in a cell-free transcription-translation system; (2) a relatively uniform and stable emulsion must be generated; (3) the MTases need to be active within emulsion droplets (e.g., some surfactants may inhibit the methyltransferase activity); and (4) a contamination by RNases must be avoided (Lee et al. 2002; Cohen et al. 2002; Tawfik

and Griffiths 1998). Nonetheless, certain precautions may be taken to enhance the probability of success; for instance, it is highly recommended to perform a "model selection" before performing the actual library selection by subjecting to selection a mixture of genes encoding an active MTase in a large excess of genes that encode a control protein (typically, at ratios of 1:100 or 1:1000). Generating a relatively uniform and stable emulsion relates to the quality of stirring and specifically to the magnetic bar spinning freely at the base of the CryoVial tube before adding the aqueous phase (for more details, see Williams et al. (2006)). Finally, RNAse inhibitors can be added to the cell-free system to prevent RNA degradation (Miller et al. 2006).

2.3 Alternative Engineering Approaches

One alternative strategy for MTase engineering involves the use of split enzymes. The bacterial M.SssI was evolved in this manner to yield an MTase, whose target specificity is dictated by a zinc finger DNA-binding domain. The wild-type MTase was split into two polypeptides that on their own were unable to reassemble into an active enzyme. Both M.SssI fragments were fused to a zinc finger domain that drove the DNA binding to CpG repeat sites as well as the assembly of both fragments to yield an active methyltransferase. This strategy led to the discovery of M.SssI variants with improved activity toward the evolved target sites and no activity on non-targeted sites (Chaikind and Ostermeier 2014). While this strategy was applied by expression and methylation in *E. coli*, it can in principle also be applied in vitro. A similar engineering was reported earlier, whereby an MTase was fused to a DNA-binding zinc finger domain with no subsequent optimization by directed evolution (Nomura and Barbas 2007).

Other potential selection approaches might be based on either in vivo or in vitro MTase expression and methylation. Selection, however, does not need to be based on restriction digest. Alternative approaches may involve a pull-down of methylated DNA using an anti-5mC antibody (methylated DNA immunoprecipitation or MeDIP (Mohn et al. 2009; Borgel et al. 2012; Zhao et al. 2014; Brebi-Mieville et al. 2012)) or other proteins that bind to methylated-CpG sequences (e.g., methylated-CpG island recovery assay or MIRA) (Jin et al. 2010). The main advantage of such strategy is that it would enable to select for methylation of CpG sites regardless of their specific sequence context. However, the sensitivity of the pull-down methods might still be too low for efficient library selection (Edelheit et al. 2013; Mohn et al. 2009).

3 Targets for Engineering

3.1 Enhanced Soluble Expression and Stability

In general, MTases are known to express poorly, and this seems to be so with both bacterial and eukaryotic enzymes. Poor expression is typically due to a low stability, namely, due to a high tendency for misfolding and aggregation (Daujotyte et al. 2003). Thus, upon overexpression, most MTases are found as insoluble aggregates.

Detailed structure-function studies, and engineering, demand reasonable expression levels and storage stability. Hence, some sequence optimization with the aim of boosting expression levels and stability is often required.

There seems to be specific solutions for certain MTases that relate to the end polishing or C-terminal truncation. Daujotyte et al. reported an attempt to shorten the C-terminus of M.HhaI to increase its solubility. The last four amino acids were substituted with a glycine. The truncated form was as active as the wild-type enzyme, while exhibiting a higher solubility (>0.35 mM). Nuclear magnetic resonance (NMR) showed a soluble, monomeric, well-folded enzyme (Daujotyte et al. 2003). Efforts to increase solubility and activity have also been described for eukaryotic Dnmts. The catalytic domain of Dnmt3a and Dnmt3b was truncated (18 amino acids at the N-terminus); the expression and activity of the catalytic domains and truncated catalytic domains were compared to the full-length wild-type enzymes (Gowher and Jeltsch 2002).

Another way of stabilizing mammalian Dnmts involves their fusion to other domains with which they are associated in vivo. Dnmt3L is a member of the Dnmt family with clear sequence homology to Dnmt3a and Dnmt3b, but with no catalytic activity. Dnmt3L is known to interact directly with the catalytic domains of Dnmt3a and Dnmt3b and stimulate their activity both in vivo and in vitro (Gowher et al. 2005). A single-chain variant in which Dnmt3a was fused to Dnmt3L via a long linker resulted in a largely aggregated protein. However, the soluble, purified single-chain protein showed about tenfold higher DNA methylation activity in vitro compared to the catalytic domain of Dnmt3a and slightly higher activity than Dnmt3a and Dnmt3L as separate domains (Siddique et al. 2013).

A general and relatively easy to implement method is the introduction of stabilizing consensus/ancestor mutations. This method can be applied as long as ≥10 orthologous sequences are available. Such mutations are also highly valuable when included in gene libraries for directed evolution (Rockah-Shmuel et al. 2014; Rockah-Shmuel and Tawfik 2012). The consensus/ancestor approach was successfully implemented to increase M. HaeIII's stability and evolvability (i.e., to boost the enzyme's ability to accept a wider range of mutations) (Rockah-Shmuel and Tawfik, 2012).

Briefly, the identification of stabilizing consensus/ancestor mutations involves the following steps: (1) creating an alignment of orthologous MTases with ≥40 % amino acids identity (REBASE is a readily available source of such sequences – http://rebase.neb.com/rebase/rebase.html) and (2) identification of positions whereby the amino acid in the target MTase deviates from the amino acid that dominates the alignment (the consensus amino acid). There are several considerations regarding the above steps. The alignment would ideally include a range of sequences with equally variable level of divergence. To this end, redundant sequences can be filtered using CD-HIT (http://weizhongli-lab.org/cd-hit/). In a typical MTase catalytic domain (~300-amino-acid protein), about 10 or so consensus mutations should be available, mostly located at surface positions. It is also important to exclude mutations in the DNA or AdoMet binding sites, although if adequately implemented, the consensus analysis should not identify such mutations (these positions are typically 100 % conserved) (see also (Goldsmith and Tawfik 2013)).

The consensus approach was pioneered in the bacterial MTase M. HaeIII (Rockah-Shmuel and Tawfik 2012). About 55 nonredundant orthologous sequences were identified and aligned (redundancy was filtered to <95 %). Eight positions were identified, where M. HaeIII clearly deviates from the consensus amino acid, which are primarily located on the surface of the enzyme (Table 1). Eight single mutants were constructed where the wild-type amino acid was exchanged into the

Table 1 The identification of consensus mutations in M. HaeIII

	Position	18	23	26	77	104	115	181	283
Identity	M. HaeIII	Q	R	C	G	I	M	F	V
	Consensus	E	E	A	A	K	L	L	I
74 %	M. Csp10ORF1501P	E	K	T	A	K			I
72 %	M. SmoLORF160P	K	D	A	A	K			I
72 %	M. LinAORF636P	K	E	A	A	K			I
72 %	M. Sin395ORF124P	E	N	A	A	K			I
72 %	M. Ava51170ORF1213P		E	A	A	K			I
72 %	M. Fma29328ORF72P	I	N	V	A	K		L	I
71 %	M. EspBORFAP	E	E	A	A	K		L	
70 %	M. Pmi33270ORF1413P	E	N	V	A	R			I
70 %	M. TasMCE3ORF1548P	E	Q	T	A	M		L	I
70 %	M. LinDORF835P	K	E	V	A	K			I
69 %	M. FnuDI	E	E	V	A	K	L	L	I
68 %	M. HbiCORF700P	E	K	E	A	L	L	L	L
67 %	M. AorOORF1855P	E	K	A	A	K		L	
67 %	M1. Bsa18170ORF1534P	H	K	I	A	L	L	L	I
67 %	M2. BspD20ORF1123P	H		T	A	L	L	L	
66 %	M2. BovSDORF2192P	H	K	T	A	L	L	L	
65 %	M. Npo43768ORF624P	E	E	V	A	K	L	L	
65 %	M. Pme18ORF425P	H		V	A	L	L	L	
64 %	M. Hha21621ORF353P	E	Q	V	A	K	L	L	I
64 %	M. Mca72ORF5914P	K	K	V	A	K	L	L	I
64 %	M. Hpy135ORF4205P	E	K	V	A	K	L	L	I
64 %	M. Pve319ORF1285P	H		V	A	L	L	L	
64 %	M. Pda43325ORF1059P	E	E	V	A	K	L	L	
64 %	M. Pru23ORF938P	E	E	V	A	K	L	L	
64 %	M. Gva14018ORFAP	E	E	V	A	K	L	L	I
63 %	M. HacSORF1213P	E	K	I	A	K	L	L	I
63 %	M. Smu29453ORF109P	E	E	V	A	K	L	L	L
63 %	M. NgoPII	E	E	A	A	K	L	L	I
62 %	M. EreORF628P	E	N	V	A	K	L	L	
60 %	M. Aur25976ORF950P	E	E	V	A	K	L	L	I
60 %	M3. BovSDORF2192P	E	D	W	A	L	L	L	I
59 %	M. Pue603ORF1201P	H	T	L	A	K	L	L	

(continued)

Table 1 (continued)

	Position	18	23	26	77	104	115	181	283
Identity	M. HaeIII	Q	R	C	G	I	M	F	V
	Consensus	E	E	A	A	K	L	L	I
58 %	M. Fco49512ORF7650P		N	W	A	K	L	L	I
57 %	M. Sde6946ORF1722P	E	K	W	A	L	L	L	I
57 %	M. Ftu200ORFAP	E	D	W	A	S	L	L	I
57 %	M. UbaSCBORF14710P		N	W	A	K	L	L	I
57 %	M. Mcu35242ORF1277P	E	N	V	A	K	L	L	I
57 %	M. DspBAVORF79P	E	N	W	A	S	L	L	I
56 %	M. AhaBGORF3490P		N	F		S	L	L	I
55 %	M. PalDORF1485P	E	N	W	A	K	E	L	I
54 %	M. MvaFGP2ORF2206P	R	N	W	A	V	L	L	
54 %	M. DthLORF402P	E	S	W	A	Y	L	L	I
53 %	M. MthTI	T	N	F	A	L	V	L	I
52 %	M. Osp6506ORF3370020P	S	N	W	A	L	L	L	I
50 %	M. Mch7420ORF2200P	R	N	W	A	R	L	L	I
50 %	M1. BovSDORF2192P	E	E	W		K	I	L	I
49 %	M. Pac4ORF808P	R	G	F	A	K	L	L	I
49 %	M1. BspFAAORF965P	E	E	W		K	I	L	I
43 %	M. EsaSS966P	–	–	–	A	K	L	L	I
40 %	M. EsaSS1545P	K	K	W	A	N	L	L	–
38 %	M. UcaBDORFAP	E	H	Y	A	F	L	L	I
38 %	M. Eba213ORF3267P	E	K	W	S	L	L	M	I
38 %	M. BhaII	E	K	W	A	L	L	L	I
37 %	M. CthORF2320P	E	K	W	A	Y	L	L	I
34 %	M. EsaSS1001P	E	K	W	A	K		L	–

The table lists eight positions in which consensus mutations were identified. Highlighted in gray are the four mutations that were identified as stabilizing by screening. Identity denotes the overall amino acid identity compared to M. HaeIII. All sequences that deviate from M. HaeIII at these eight positions are listed (adopted from the supplementary material of Rockah-Shmuel and Tawfik (2012))

consensus one. The levels of soluble enzyme and its activity were tested in crude lysates following overexpression in *E. coli*. In this manner, four stabilizing mutations were found that were subsequently introduced into the wild-type M. HaeIII (C26A, I104K, M115L, and F181L). The stabilized variant gave ~fivefold increase in the yield of the soluble and active enzyme (Fig. 2) and was subsequently used as the starting point for directed evolution of M. HaeIII toward new DNA methylation targets (Rockah-Shmuel et al. 2015; Rockah-Shmuel et al. 2013; Bloom et al. 2006).

Apart from the consensus, several other criteria can be applied to identify stabilizing mutations, such as high side-chain mobility, which is indicated by high B-factors, in cases where a crystal structure is available (Augustyniak et al. 2012). Indeed, when high-resolution crystal structures are available, computational design methods can be applied to significantly boost stability and expression (Wijma et al. 2014).

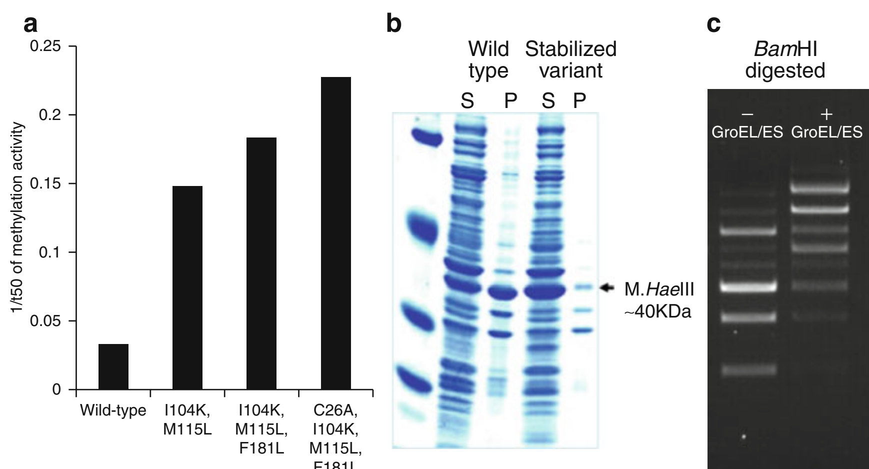


Fig. 2 Increasing the soluble expression of M. HaeIII by the introduction of consensus mutations and co-expression of the GroEL/ES chaperone (adopted from the supplementary material of Rockah-Shmuel and Tawfik (2012)). (**a**) Methylation activity of consensus M. HaeIII variants. Increases in methylation activities in crude lysates were plotted as $1/(t_{50})$, whereby t_{50} is the time required to methylate 50 % of the DNA substrate. Methylation was assayed using an ELISA format with a DIG-Biotin DNA substrate. (**b**) SDS-PAGE of *E. coli* supernatant (S) and pellet (P) of over-expressed wild-type and consensus mutants of M. HaeIII. For the wild-type M. HaeIII, <40 % of the enzyme is soluble (i.e., in the supernatant), whereas the stabilized variant carrying four consensus mutations (C26A, I104K, M115L, and F181L) exhibits a much higher fraction of soluble protein (>90 %). (**c**) The effect of GroEL/ES chaperonin overexpression on M. HaeIII methylation activity. The evolved M. BamHI-like B1 variant (G22) shows low plasmid protection levels when transformed to the MC1051 *E. coli* strain. Induction of a GroEL/ES expression plasmid (by adding arabinose to the growth medium) increased the level of methylation as manifested in a significantly lower fraction of the digested plasmid

A clear caveat associated with the introduction of stabilizing mutations is that they may alter the MTase's properties. This is not so much of a concern with bacterial R/M MTases but is a major caveat with respect to eukaryotic Dnmts involved in epigenetics. Thus, an alternative way of obtaining higher yields of soluble, active enzyme with no sequence alterations is by co-expression of chaperones, most commonly GroEL/ES, as exemplified in Fig. 2c. GroEL/ES co-expression may also be used to facilitate the evolution of enzyme variants bearing new functions (Tokuriki and Tawfik 2009a).

3.2 Alteration of the DNA Target Sequence

A major effort pursued during the last two decades has been to engineer MTases with new DNA target specificities (Table 2). The initial attempts were not so successful and reported primarily relaxed specificity. More recently, the engineering of bacterial MTases that methylate new sites has been reported.

The initial attempts resulted in the expansion of "star" activities (methylation of sites that differ from the original target site by one base) (Cohen et al. 2004;

Table 2 Engineered and evolved MTase variants with altered or new DNA target specificities

Enzyme	Type	Original DNA target	New DNA target	Change in specificity (fold)[a]	Change in rate (fold)[b]	Refs.
M. HaeIII	C5	GG**C**C	**A**GCC	>10	670	Cohen et al. (2004)
EcoDam	N6	GAT**C**	GAT**T**	10	1600	Chahar et al. (2010)
M. SinI	C5	GG (**A/T**) CC	GG (**G/C**) CC	Similar to wild type	20	Timar et al. (2004)
M. HhaI	C5	GCGC	GCG	10-fold	n.r.	Gerasimaite et al. (2009)
M. HaeIII	C5	G**G**CC	GG**AT**CC	$>10^4$	$>10^5$	Rockah-Shmuel and Tawfik (2012)
M. HaeIII	C5	G**G**CC	GG (**A/T**) CC	660	$>10^2$	Rockah-Shmuel and Tawfik (2012)
M. HaeIII	C5	G**G**CC	G**C**G**C**	$>10^4$	$>10^6$	Rockah-Shmuel and Tawfik (2012)

n.r. not reported

[a]The change in ratios of k_{cat}/K_M values for the new target site versus the original site in the evolved variant compared to wild type

[b]The ratio of k_{cat}/K_M values for the target site in the evolved variant versus wild type

Chahar et al. 2010). For instance, M. HaeIII was evolved by in vitro compartmentalization to methylate **A**GCC sites in addition to the original **G**GCC sites. Overall, these mutants exhibited a broad specificity, including the methylation of non-palindromic sites (Cohen et al. 2004). Curiously, these mutants also exhibited higher catalytic efficiency compared to the wild-type enzyme, not just toward the evolved star sites but also toward the original site, GGCC. This suggests the existence of a rate-accuracy trade-off, whereby the fidelity of MTases comes at the expense of their methylation rates (Tawfik 2014).

A similar example regards a change of the target specificity of EcoDam from GATC to GATT by directed evolution (Chahar et al. 2010). Notably, the enzyme and its DNA target substrate were coevolved by using a library of enzyme variants as well as of different DNA methylation sites (GANN). The evolved enzyme variants exhibited a 1600-fold change in specificity for GATT methylation.

However, with time, it became doubtful whether such star-methylating variants comprise "on-pathway" intermediates to new palindromic targets. Indeed, our attempts to further evolve such variants to methylate new palindromic sites failed consistently. For example, we made several attempts to take M. HaeIII variants evolved for AGCC and evolved them further to methylate AGCT sites, but none of these attempts were successful (Bershtein et al. unpublished results). Other examples along the same vein include the in vitro evolution of M. SinI for a relaxed GGNCC recognition specificity (Timar et al. 2004). The resulting variants exhibited higher methylation activity toward GG (G/C) CC, unlike the wild type that primarily methylates GG (A/T) CC. Another example regards the "truncation" of one base from the original methylation site for M. HhaI, thus switching specificity from GCGC to GCG (Gerasimaite et al. 2009).

A crucial insight regarding how new DNA methylation specificities evolve came from the study of M. EcoRV – a bacterial adenine-N6 MTase that methylates GATATC. EcoDam is its homologous bacterial adenine-N6 MTase that methylates GATC. It appears that M. EcoRV bends its longer DNA target site, suggesting an indirect readout of the TA inner sequence and the evolution of new target specificities by insertion-deletion of existing target sites (Jurkowski et al. 2007). This evolutionary mechanism was directly supported by the directed evolution of a bacterial C5-MTase, M. HaeIII (Rockah-Shmuel and Tawfik 2012). Thus, new enzyme variants readily evolved that methylate target sites that comprise an extension of the original GGCC, e.g., GG (A/T) CC and GG (CG) CC. Interestingly, the evolved enzyme variants also showed methylation of novel sites that are not methylated by the wild type, nor that were selected for, including new palindromic sites such as GCGC or GGATCC. This experiment indicated that evolutionary intermediates possess a "generalist" character, namely, a broad methylation specificity that can be exploited to provide starting points for a range of new specificities that are not available in a given wild-type enzyme. The experiment also validated the mechanism of target site expansion or shrinkage as the means by which new DNA target specificities evolve in nature.

3.3 Engineering for Usage of Alternative Cofactors

Engineering for alternative substrates typically relies on promiscuity, namely, on enzymes accepting substrates or cofactors other than their native ones. The engineering of MTases toward new cofactors, therefore, involves the application of engineered, synthetic analogs of the natural cofactor, S-adenosyl-L-methionine (AdoMet), that are accepted, even with low efficiency, by the natural wild-type enzymes. Subsequently, MTase mutants can be selected to accept the artificial cofactor with high efficiency and selectivity, while excluding the natural cofactor.

AdoMet can be divided into four structural elements, with the first one, the transferred methyl group (1, Fig. 3) being the most critical one. In fact, MTases (and other methyltransferases; see, e.g., Bothwell et al. (2012)) appear to exhibit a considerable promiscuity with respect to the methyl group. This promiscuity enables the use of AdoMet analogs (hereafter–*1*-AdoMet), thereby leading to the alkylation of DNA bases with groups other than methyl. Such *1*-AdoMets can be tractable by virtue of having, for example, biotin or azide or alkyne groups for click chemistry (Vranken et al. 2014; Lukinavicius et al. 2013). Methyl is effectively the smallest carbon-based moiety, and thus, by default, *1*-AdoMet analogs result in the modification of DNA bases with bulkier groups, thus resulting in biological effects that may differ from methylation. The limits of the MTase's acceptance of the *1*-AdoMet analogs appear to be primarily electronic rather than steric. Specifically, extended 1' groups containing various functional moieties (e.g., a primary amine) are accepted by bacterial MTases as long as they contain a double or triple bond at the β-carbon (relative to the sulfonium group) (Dalhoff et al. 2006).

While DNA methyltransferases (and other methyltransferases) can accept *1*-AdoMet analogs, the rates of alkyl-DNA transfer are typically much lower than

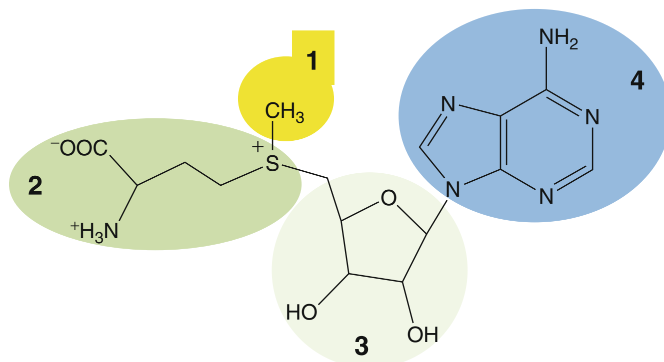


Fig. 3 Chemical structure of S-adenosyl-L-methionine. Highlighted in different colors are its four structural elements: *1* the transferred methyl group, *2* the S-homocysteine group, *3* the ribose, and *4* the adenine

with AdoMet. Thus, in a living cell, where the original cofactor (AdoMet) is present usually at relatively high concentrations (e.g., ~0.1–0.2 mM in *E. coli* (Bennett et al. 2009)), a modified cofactor may not be able to compete with it. By exploring mutations in residues within and close to the MTase methyl-binding pocket, enzyme variants of the bacterial C5 MTase M. HhaI were selected that preferentially accept *1*-AdoMet analogs (Lukinavicius et al. 2012). These engineered variants show marked increases in incorporating larger 1'-alkyl groups containing four carbons, such as but-2-ynyl (–CH_2-CΞC-CH_3), and parallel decreases of methyl incorporation (with AdoMet as cofactor). The high similarity between bacterial C5-MTases means that the same or similar mutations may yield higher activity with 1'-modified AdoMet analogs in other enzymes as well (Lukinavicius et al. 2012; Dalhoff et al. 2006). The mutated M. HhaI containing up to three mutations, including Q82A, Y254S, and N304A, exhibited ~tenfold preference for alkyl-*1*-AdoMet analogs relative to AdoMet. While modest, achieving this level of selectivity is far from being trivial. Because of the huge energetic penalty of steric clashes, in general, discriminating against a smaller substrate is far more challenging than against a larger one (Tawfik 2014). The engineered mutants were also capable of alkylating DNA with a long linker (ten carbons) that also contained a terminal amine group. However, the rates and selectivities were significantly lower than with the C4 group.

In the context of alternative DNA modifications, it is worth noting that MTases can also act as azidonucleosidyl transferases. This involves the replacement of both the S-homocysteine and methyl groups of AdoMet (1, 2, Fig. 3) with an aziridine group. The bacterial A6-MTase TaqI was shown to accept this cofactor and thus covalently modify DNA with an adenosyl rather than a methyl group. The adenosyl group of these AdoMet analogs (3, Fig. 3) can be further modified to carry a fluorescent group or biotin, thus enabling to modify DNA with a tractable marker (Pignot et al. 1998; Pljevaljcic et al. 2003; Comstock and Rajski 2005a, b).

To our knowledge, there are no reported attempts of engineering or evolving MTases to preferentially accept the aziridine analog of AdoMet. Such engineering will demand changes primarily in the amino acids involved in the binding of the S-homocysteine group (2, Fig. 3). Indeed, so far, the engineering toward alternative cofactors has been limited to the above-described replacements of the methyl group.

Nonetheless, modifications of the other elements of AdoMet (2–4; Fig. 3) should be possible.

Beyond the methyl group, the cofactor promiscuity of MTases has not been systematically explored. One study used as a probe S-adenosyl-L-homocysteine, the product of AdoMet after methyl group transfer, which binds MTases with an affinity that is similar to AdoMet. It showed that modifications in any of the elements *2–4*, such as the amino or carboxyl groups of the S-homocysteine moiety (moiety 2), the hydroxyls of the ribose (3), or the base (4, Fig. 3), generally result in >1,000-fold decreases in binding affinity (Cohen et al. 2005). Nonetheless, given the high-binding affinity of AdoMet, certain analogs with some of the abovementioned modifications might still serve as cofactors. Given that the AdoMet binding pockets of MTases are structurally well characterized, it is also likely that MTases could be readily engineered to accept such AdoMet analogs. For example, binding of the ribose's hydroxyls is mediated by a bidentate interaction with a canonical Glu that appears in all AdoMet-dependent methyltransferases, including MTases (but excluding radical-AdoMet enzymes). Mutations in this Glu result in ≥300-fold decrease in AdoMet binding (Laurino et al. 2016). Ultimately, synthetic modifications of the cofactor, including the replacements of the methyl group, and mutations in the AdoMet binding residues may enable complete orthogonality – namely, that, on the one hand, engineered MTase variants efficiently modify DNA with the alternative cofactor and do not accept AdoMet and, on the other hand, that the alterative cofactor is not accepted by the natural MTases.

It is also worth noting that although the above-described works made use of AdoMet analogs generated by organic synthesis, there exists an appealing alternative of using the enzyme responsible for the biosynthesis of AdoMet, namely, AdoMet synthase. This enzyme condenses ATP and methionine to yield AdoMet. The promiscuity of AdoMet synthase has been explored in quite detail and in AdoMet synthases from different organisms. It appears that both analogs of methionine and of ATP, including dATP, are accepted, in some cases with surprisingly high rates. This promiscuity enables the synthesis of AdoMet analogs with modifications of any of AdoMet's four elements (Lu and Markham 2002) and thus opens the potential for generating new engineered MTases working with alternative cofactors which can be synthesized within a living cell. The latter provides a distinct advantage, given that AdoMet was reported to have low permeability to cells (Wang et al. 2014; Lin et al. 2001).

4 Concluding Remarks

The engineering of MTases may open the road to new ways of manipulating DNA in a unique and selective manner. Obtaining new methylation target specificities is one option, as is the modification of DNA with tractable groups instead of methyl. The latter is mediated by analogs of AdoMet, in which the methyl group of AdoMet had been replaced. Modified cofactors may also serve in chemical epigenetics, meaning in generating Dnmts whose methylation activity is triggered by an artificial

cofactor. However, while the prospects for MTase engineering have been demonstrated, there remain key challenges that need to be overcome. Overall, the catalytic efficiency of engineered and evolved MTases is relatively low, and foremost, their selectivity is low. Indeed, the challenge of obtaining real orthogonality is yet to be met (i.e., obtaining enzyme variants, whose activity with the original DNA target or with AdoMet is effectively nil, while their activity with the new target or cofactor is comparable to wild type with its original target and AdoMet).

Another key challenge that is of fundamental interest regards the coevolution of R/M pairs. The R/M system is composed of two highly selective enzymes, a methyltransferase and a restriction endonuclease, both acting on the same DNA target. Evolutionary, this system is extremely interesting, because of its “cooperativity” (Mruk and Blumenthal 2008; Kobayashi 2001). In principle, a restriction endonuclease without a cognate MTase is toxic. However, an MTase provides no advantage unless a cognate endonuclease exists. This hen-egg cycle may be resolved by promiscuous activities of both MTases (as described above) and restriction endonucleases (Sapienza et al. 2005; Samuelson et al. 2006). However, the manners by which overlapping promiscuous activities appear, and how they behave under selection, are unknown. The reconstruction of an evolutionary trajectory leading to a new MTase-endonuclease pair may therefore provide important insights into the evolution of R/M systems in particular and the coevolution of toxin-antitoxin systems in general.

References

Augustyniak W, Brzezinska AA, Pijning T, Wienk H, Boelens R, Dijkstra BW, Reetz MT. Biophysical characterization of mutants of Bacillus subtilis lipase evolved for thermostability: factors contributing to increased activity retention. Protein Sci. 2012;21(4):487–97. doi:10.1002/pro.2031.

Bennett BD, Kimball EH, Gao M, Osterhout R, Van Dien SJ, Rabinowitz JD. Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. Nat Chem Biol. 2009;5(8):593–9. doi:10.1038/nchembio.186.

Bernath K, Magdassi S, Tawfik DS. In vitro compartmentalization (IVC): a high-throughput screening technology using emulsions and FACS. Discov Med. 2004;4(20):49–53.

Bernath K, Magdassi S, Tawfik DS. Directed evolution of protein inhibitors of DNA-nucleases by in vitro compartmentalization (IVC) and nano-droplet delivery. J Mol Biol. 2005;345(5):1015–26. doi:10.1016/j.jmb.2004.11.017.

Bloom JD, Labthavikul ST, Otey CR, Arnold FH. Protein stability promotes evolvability. Proc Natl Acad Sci U S A. 2006;103(15):5869–74. doi:10.1073/pnas.0510098103.

Borgel J, Guibert S, Weber M. Methylated DNA immunoprecipitation (MeDIP) from low amounts of cells. Methods Mol Biol. 2012;925:149–58. doi:10.1007/978-1-62703-011-3_9.

Bothwell IR, Islam K, Chen Y, Zheng W, Blum G, Deng H, Luo M. Se-adenosyl-L-selenomethionine cofactor analogue as a reporter of protein methylation. J Am Chem Soc. 2012;134(36):14905–12. doi:10.1021/ja304782r.

Brebi-Mieville P, Ili-Gangas C, Leal-Rojas P, Noordhuis MG, Soudry E, Perez J, Roa JC, Sidransky D, Guerrero-Preston R. Clinical and public health research using methylated DNA immunoprecipitation (MeDIP): a comparison of commercially available kits to examine differential DNA methylation across the genome. Epigenetics. 2012;7(1):106–12. doi:10.4161/epi.7.1.18647.

Chahar S, Elsawy H, Ragozin S, Jeltsch A. Changing the DNA recognition specificity of the EcoDam DNA-(adenine-N6)-methyltransferase by directed evolution. J Mol Biol. 2010;395(1):79–88. doi:10.1016/j.jmb.2009.09.027.

Chaikind B, Ostermeier M. Directed evolution of improved zinc finger methyltransferases. PLoS One. 2014;9(5):e96931. doi:10.1371/journal.pone.0096931.

Cohen HM, Griffiths AD, Tawfik DS, Loakes D. Determinants of cofactor binding to DNA methyltransferases: insights from a systematic series of structural variants of S-adenosylhomocysteine. Org Biomol Chem. 2005;3(1):152–61. doi:10.1039/b415446k.

Cohen HM, Tawfik DS, Griffiths AD. Promiscuous methylation of non-canonical DNA sites by HaeIII methyltransferase. Nucleic Acids Res. 2002;30(17):3880–5.

Cohen HM, Tawfik DS, Griffiths AD. Altering the sequence specificity of HaeIII methyltransferase by directed evolution using in vitro compartmentalization. Protein Eng Des Sel. 2004;17(1): 3–11. doi:10.1093/protein/gzh001.

Comstock LR, Rajski SR. Conversion of DNA methyltransferases into azidonucleosidyl transferases via synthetic cofactors. Nucleic Acids Res. 2005a;33(5):1644–52. doi:10.1093/nar/gki306.

Comstock LR, Rajski SR. Methyltransferase-directed DNA strand scission. J Am Chem Soc. 2005b;127(41):14136–7. doi:10.1021/ja054128y.

Dalby PA. Strategy and success for the directed evolution of enzymes. Curr Opin Struct Biol. 2011;21(4):473–80. doi:10.1016/j.sbi.2011.05.003.

Dalhoff C, Lukinavicius G, Klimasauskas S, Weinhold E. Direct transfer of extended groups from synthetic cofactors by DNA methyltransferases. Nat Chem Biol. 2006;2(1):31–2. doi:10.1038/nchembio754.

Daujotyte D, Vilkaitis G, Manelyte L, Skalicky J, Szyperski T, Klimasauskas S. Solubility engineering of the HhaI methyltransferase. Protein Eng. 2003;16(4):295–301.

Edelheit S, Schwartz S, Mumbach MR, Wurtzel O, Sorek R. Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. PLoS Genet. 2013;9(6):e1003602. doi:10.1371/journal.pgen.1003602.

Friedrich T, Roth M, Helm-Kruse S, Jeltsch A. Functional mapping of the EcoRV DNA methyltransferase by random mutagenesis and screening for catalytically inactive mutants. Biol Chem. 1998;379(4–5):475–80.

Gerasimaite R, Vilkaitis G, Klimasauskas S. A directed evolution design of a GCG-specific DNA hemimethylase. Nucleic Acids Res. 2009;37(21):7332–41. doi:10.1093/nar/gkp772.

Goldsmith M, Tawfik DS. Enzyme engineering by targeted libraries. Methods Enzymol. 2013;523:257–83. doi:10.1016/B978-0-12-394292-0.00012-6.

Gowher H, Jeltsch A. Molecular enzymology of the catalytic domains of the Dnmt3a and Dnmt3b DNA methyltransferases. J Biol Chem. 2002;277(23):20409–14. doi:10.1074/jbc.M202148200.

Gowher H, Liebert K, Hermann A, Xu G, Jeltsch A. Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. J Biol Chem. 2005;280(14):13341–8. doi:10.1074/jbc.M413412200.

Griffiths AD, Tawfik DS. Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. EMBO J. 2003;22(1):24–35. doi:10.1093/emboj/cdg014.

Hart DJ, Waldo GS. Library methods for structural biology of challenging proteins and their complexes. Curr Opin Struct Biol. 2013;23(3):403–8. doi:10.1016/j.sbi.2013.03.004.

Hughes JA, Brown LR, Ferro AJ. Expression of the cloned coliphage T3 S-adenosylmethionine hydrolase gene inhibits DNA methylation and polyamine biosynthesis in Escherichia coli. J Bacteriol. 1987;169(8):3625–32.

Jeltsch A, Sobotta T, Pingoud A. Structure prediction of the EcoRV DNA methyltransferase based on mutant profiling, secondary structure analysis, comparison with known structures of methyltransferases and isolation of catalytically inactive single mutants. Protein Eng. 1996;9(5):413–23.

Jin SG, Kadam S, Pfeifer GP. Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. Nucleic Acids Res. 2010;38(11):e125. doi:10.1093/nar/gkq223.

Jochens H, Bornscheuer UT. Natural diversity to guide focused directed evolution. Chembiochem. 2010;11(13):1861–6. doi:10.1002/cbic.201000284.

Jurkowska RZ, Siddique AN, Jurkowski TP, Jeltsch A. Approaches to enzyme and substrate design of the murine Dnmt3a DNA methyltransferase. Chembiochem. 2011;12(10):1589–94. doi:10.1002/cbic.201000673.

Jurkowski TP, Anspach N, Kulishova L, Nellen W, Jeltsch A. The M.EcoRV DNA-(adenine N6)-methyltransferase uses DNA bending for recognition of an expanded EcoDam recognition site. J Biol Chem. 2007;282(51):36942–52. doi:10.1074/jbc.M706933200.

Kobayashi I. Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. Nucleic Acids Res. 2001;29(18):3742–56.

LaMonte BL, Hughes JA. In vivo hydrolysis of S-adenosylmethionine induces the met regulon of Escherichia coli. Microbiology. 2006;152(Pt 5):1451–9. doi:10.1099/mic.0.28489-0.

Laurino P, Toth-Petroczy A, Meana-Pañeda R, Lin W, Truhlar DG, Tawfik DS. An ancient fingerprint indicates the common ancestry of Rossmann fold enzymes utilizing different ribose based cofactors. PloS Biol. 2016;14(3):e1002396.

Lee YF, Tawfik DS, Griffiths AD. Investigating the target recognition of DNA cytosine-5 methyltransferase HhaI by library selection using in vitro compartmentalisation. Nucleic Acids Res. 2002;30(22):4937–44.

Lin Q, Jiang F, Schultz PG, Gray NS. Design of allele-specific protein methyltransferase inhibitors. J Am Chem Soc. 2001;123(47):11608–13.

Lu ZJ, Markham GD. Enzymatic properties of S-adenosylmethionine synthetase from the archaeon Methanococcus jannaschii. J Biol Chem. 2002;277(19):16624–31. doi:10.1074/jbc.M110456200.

Lukinavicius G, Lapinaite A, Urbanaviciute G, Gerasimaite R, Klimasauskas S. Engineering the DNA cytosine-5 methyltransferase reaction for sequence-specific labeling of DNA. Nucleic Acids Res. 2012;40(22):11594–602. doi:10.1093/nar/gks914.

Lukinavicius G, Tomkuviene M, Masevicius V, Klimasauskas S. Enhanced chemical stability of adomet analogues for improved methyltransferase-directed labeling of DNA. ACS Chem Biol. 2013;8(6):1134–9. doi:10.1021/cb300669x.

Miller OJ, Bernath K, Agresti JJ, Amitai G, Kelly BT, Mastrobattista E, Taly V, Magdassi S, Tawfik DS, Griffiths AD. Directed evolution by in vitro compartmentalization. Nat Methods. 2006;3(7):561–70. doi:10.1038/nmeth897.

Miyazaki K. Creating random mutagenesis libraries by megaprimer PCR of whole plasmid (MEGAWHOP). Methods Mol Biol. 2003;231:23–8. doi:10.1385/1-59259-395-X:23.

Mohn F, Weber M, Schubeler D, Roloff TC. Methylated DNA immunoprecipitation (MeDIP). Methods Mol Biol. 2009;507:55–64. doi:10.1007/978-1-59745-522-0_5.

Morley KL, Kazlauskas RJ. Improving enzyme properties: when are closer mutations better? Trends Biotechnol. 2005;23(5):231–7. doi:10.1016/j.tibtech.2005.03.005.

Mruk I, Blumenthal RM. Real-time kinetics of restriction-modification gene expression after entry into a new host cell. Nucleic Acids Res. 2008;36(8):2581–93. doi:10.1093/nar/gkn097.

Nomura W, Barbas 3rd CF. In vivo site-specific DNA methylation with a designed sequence-enabled DNA methylase. J Am Chem Soc. 2007;129(28):8676–7. doi:10.1021/ja0705588.

Pignot M, Siethoff C, Linscheid M, Weinhold E. Coupling of a Nucleoside with DNA by a Methyltransferase. Angew Chem Int Ed Engl. 1998;37(20):3.

Pljevaljcic G, Pignot M, Weinhold E. Design of a new fluorescent cofactor for DNA methyltransferases and sequence-specific labeling of DNA. J Am Chem Soc. 2003;125(12):3486–92. doi:10.1021/ja021106s.

Raleigh EA, Wilson G. Escherichia coli K-12 restricts DNA containing 5-methylcytosine. Proc Natl Acad Sci U S A. 1986;83(23):9070–4.

Rockah-Shmuel L, Tawfik DS. Evolutionary transitions to new DNA methyltransferases through target site expansion and shrinkage. Nucleic Acids Res. 2012;40(22):11627–37. doi:10.1093/nar/gks944.

Rockah-Shmuel L, Tawfik DS, Goldsmith M. Generating targeted libraries by the combinatorial incorporation of synthetic oligonucleotides during gene shuffling (ISOR). Methods Mol Biol. 2014;1179:129–37. doi:10.1007/978-1-4939-1053-3_8.

Rockah-Shmuel L, Toth-Petroczy A, Sela A, Wurtzel O, Sorek R, Tawfik DS. Correlated occurrence and bypass of frame-shifting insertion-deletions (InDels) to give functional proteins. PLoS Genet. 2013;9(10):e1003882. doi:10.1371/journal.pgen.1003882.

Rockah-Shmuel L, Toth-Petroczy A, Tawfik DS. Systematic mapping of protein mutational space by prolonged drift reveals the deleterious effects of seemingly neutral mutations. PLoS Comput Biol. 2015;11(8):e1004421. doi:10.1371/journal.pcbi.1004421.

Samuelson JC, Morgan RD, Benner JS, Claus TE, Packard SL, Xu SY. Engineering a rare-cutting restriction enzyme: genetic screening and selection of NotI variants. Nucleic Acids Res. 2006;34(3):796–805. doi:10.1093/nar/gkj483.

Sapienza PJ, Dela Torre CA, McCoy WH, Jana SV, Jen-Jacobson L. Thermodynamic and kinetic basis for the relaxed DNA sequence specificity of "promiscuous" mutant EcoRI endonucleases. J Mol Biol. 2005;348(2):307–24. doi:10.1016/j.jmb.2005.02.051.

Siddique AN, Nunna S, Rajavelu A, Zhang Y, Jurkowska RZ, Reinhardt R, Rots MG, Ragozin S, Jurkowski TP, Jeltsch A. Targeted methylation and gene silencing of VEGF-A in human cells by using a designed Dnmt3a-Dnmt3L single-chain fusion protein with increased DNA methylation activity. J Mol Biol. 2013;425(3):479–91. doi:10.1016/j.jmb.2012.11.038.

Stemmer WP. Rapid evolution of a protein in vitro by DNA shuffling. Nature. 1994;370(6488):389–91. doi:10.1038/370389a0.

Szomolanyi E, Kiss A, Venetianer P. Cloning the modification methylase gene of Bacillus sphaericus R in Escherichia coli. Gene. 1980;10(3):219–25.

Tawfik DS. Accuracy-rate tradeoffs: how do enzymes meet demands of selectivity and catalytic efficiency? Curr Opin Chem Biol. 2014;21:73–80. doi:10.1016/j.cbpa.2014.05.008.

Tawfik DS, Griffiths AD. Man-made cell-like compartments for molecular evolution. Nat Biotechnol. 1998;16(7):652–6. doi:10.1038/nbt0798-652.

Timar E, Groma G, Kiss A, Venetianer P. Changing the recognition specificity of a DNA-methyltransferase by in vitro evolution. Nucleic Acids Res. 2004;32(13):3898–903. doi:10.1093/nar/gkh724.

Tokuriki N, Tawfik DS. Chaperonin overexpression promotes genetic variation and enzyme evolution. Nature. 2009a;459(7247):668–73. doi:10.1038/nature08009.

Tokuriki N, Tawfik DS. Stability effects of mutations and protein evolvability. Curr Opin Struct Biol. 2009b;19(5):596–604. doi:10.1016/j.sbi.2009.08.003.

Ulrich A, Andersen KR, Schwartz TU. Exponential megapriming PCR (EMP) cloning--seamless DNA insertion into any target plasmid without sequence constraints. PLoS One. 2012;7(12):e53360. doi:10.1371/journal.pone.0053360.

Vranken C, Deen J, Dirix L, Stakenborg T, Dehaen W, Leen V, Hofkens J, Neely RK. Super-resolution optical DNA Mapping via DNA methyltransferase-directed click chemistry. Nucleic Acids Res. 2014;42(7):e50. doi:10.1093/nar/gkt1406.

Wang R, Zheng W, Luo M. A sensitive mass spectrum assay to characterize engineered methionine adenosyltransferases with S-alkyl methionine analogues as substrates. Anal Biochem. 2014;450:11–9. doi:10.1016/j.ab.2013.12.026.

Wijma HJ, Floor RJ, Jekel PA, Baker D, Marrink SJ, Janssen DB. Computationally designed libraries for rapid enzyme stabilization. Protein Eng Des Sel. 2014;27(2):49–58. doi:10.1093/protein/gzt061.

Williams R, Peisajovich SG, Miller OJ, Magdassi S, Tawfik DS, Griffiths AD. Amplification of complex gene libraries by emulsion PCR. Nat Methods. 2006;3(7):545–50. doi:10.1038/nmeth896.

Zhao MT, Whyte JJ, Hopkins GM, Kirk MD, Prather RS. Methylated DNA immunoprecipitation and high-throughput sequencing (MeDIP-seq) using low amounts of genomic DNA. Cell Reprogram. 2014;16(3):175–84. doi:10.1089/cell.2014.0002.

DNA Labeling Using DNA Methyltransferases

Miglė Tomkuvienė, Edita Kriukienė, and Saulius Klimašauskas

Abstract

DNA methyltransferases (MTases) uniquely combine the ability to recognize and covalently modify specific target sequences in DNA using the ubiquitous cofactor S-adenosyl-L-methionine (AdoMet). Although DNA methylation plays important roles in biological signaling, the transferred methyl group is a poor reporter and is highly inert to further biocompatible derivatization. To unlock the biotechnological power of these enzymes, two major types of cofactor AdoMet analogs were developed that permit targeted MTase-directed attachment of larger moieties containing functional or reporter groups onto DNA. One such approach (named sequence-specific methyltransferase-induced labeling, SMILing) uses reactive aziridine or *N*-mustard mimics of the cofactor AdoMet, which render targeted coupling of a whole cofactor molecule to the target DNA. The second approach (methyltransferase-directed transfer of activated groups, mTAG) uses AdoMet analogs with a sulfonium-bound extended side chain replacing the methyl group, which permits MTase-directed covalent transfer of the activated side chain alone. As the enlarged cofactors are not always compatible with the active sites of native MTases, steric engineering of the active site has been employed to optimize their alkyltransferase activity. In addition to the described cofactor analogs, recently discovered atypical reactions of DNA cytosine-5 MTases involving non-cofactor-like compounds can also be exploited for targeted derivatization and labeling of DNA. Altogether, these approaches offer new powerful tools for sequence-specific covalent DNA labeling, which not only pave the way to developing a variety of useful techniques in DNA research, diagnostics, and nanotechnologies but have already proven practical utility for optical DNA mapping and epigenome studies.

M. Tomkuvienė • E. Kriukienė • S. Klimašauskas (✉)
Institute of Biotechnology, Vilnius University, Vilnius LT-10222, Lithuania
e-mail: saulius.klimasauskas@bti.vu.lt

© Springer International Publishing Switzerland 2016

A. Jeltsch, R.Z. Jurkowska (eds.), *DNA Methyltransferases - Role and Function*,
Advances in Experimental Medicine and Biology 945,
DOI 10.1007/978-3-319-43624-1_19

1 Introduction

DNA is a large linear polymer comprised of aperiodic combinations of four major types of building blocks encoding the genetic blueprint of life. Since different loci of this largely uniform biomolecule rarely contain features distinct enough to permit their chemical or physical identification among other DNA loci or other biomolecules, a key task is to furnish them with suitable reporter tags for their selective visualization and isolation from biological samples. Among the variety of enzymes involved in DNA metabolism, DNA methyltransferases (MTases) uniquely combine two useful features required for targeted labeling: recognition of a vast repertoire of specific target sequences (2–8 nt long) and covalent modification of the target site. Although targeted DNA methylation can be “read” by specific cellular proteins and thus plays important roles in biological signaling, the naturally transferred methyl group is a poor reporter and is not readily amenable for further chemical derivatization. Therefore, one strategy to unlock the biotechnological potential of these highly specific MTase enzymes is to make them transfer “pre-derivatized” (extended) versions of the methyl group. The catalytic power of AdoMet-dependent MTases to a large extent derives from their ability to bring the two substrates, the cofactor AdoMet and a target molecule, together in the right orientation. Thus, a series of synthetic analogs of the AdoMet cofactor were developed that allowed MTases to tag DNA with extended moieties, making sequence-specific MTase-directed labeling an attractive opportunity in various biotechnological applications. Two major types of cofactor analogs have been developed for MTase-catalyzed DNA labeling which permit covalent deposition of either a whole cofactor molecule or its sulfonium-bound side chain. Among the three known classes of DNA methyltransferases (cytosine-C5, adenine-N6, and cytosine-N4 MTases), the first two have been largely utilized for the attachment of various reactive groups, biotin or fluorophores to DNA. Due to the universal nature of the AdoMet cofactor for biological methylations, the approach also proved applicable for labeling other biomolecules, such as RNA (Motorin et al. 2011; Tomkuvienė et al. 2012; Plotnikova et al. 2014; Schulz et al. 2013; Holstein et al. 2014), proteins (Peters et al. 2010; Islam et al. 2011; Willnow et al. 2012; Wang et al. 2013; Hymbaugh Bergman and Comstock 2015) and small molecules (Zhang et al. 2006; Stecher et al. 2009; Lee et al. 2010; Winter et al. 2013) using appropriate MTases.

Another recently developed cofactor-independent DNA modification strategy is based on atypical reactions of DNA cytosine-C5 MTases. Upon interaction with the target cytosine, these MTases use a covalent attack to transiently generate an activated cytosine intermediate (ACI). In the absence of AdoMet or synthetic AdoMet analogs, the ACI can undergo a covalent addition of exogenous formaldehyde yielding 5-hydroxymethylcytosine (hmC). Moreover, hmC residues at the target site can be dehydroxymethylated to yield cytosine or can undergo further addition of thiols or selenols to yield the corresponding 5-chalcogenomethyl derivatives in DNA in a C5-MTase-dependent manner. These transformations open new possibilities for sequence-specific derivatization and analysis of epigenetic marks in mammalian

DNA. In the following sections, the DNA labeling approaches based on the two types of synthetic cofactor analogs and the reactions involving non-cofactor-like compounds are discussed in detail.

2 Synthetic Cofactor Analogs for MTase-Directed Modification of DNA

The first labeling strategy (named sequence-specific methyltransferase-induced labeling or SMILing) developed by the Weinhold group employed cofactor analogs in which the methionine moiety of AdoMet was synthetically replaced by an aziridine ring (*N*-adenosylaziridine cofactors) (Pignot et al. 1998). Upon reaction of *N*-adenosylaziridine with DNA in the presence of a DNA MTase, the "transfer" of an electrophilic carbon atom of the protonated aziridine ring to a nucleophilic target atom in DNA leads to ring opening, thereby turning the ring into an ethylamino linker that connects the cofactor molecule with the target nucleobase (Pignot et al. 1998) (Fig. 1). Although the attached ethylaminoadenosine moiety by itself is not a good reporter group, it can serve as a carrier to which desired chemical and reporter groups are attached. To minimize interference with proper cofactor binding in the catalytic center of a directing MTase, the selection of potential anchoring points in the adenosine moiety appears to be limited to the 6, 7, and 8 positions of the adenine ring (Pljevaljčić et al. 2004; Kunkel et al. 2015). The SMILing approach was initially developed with an aziridine AdoMet analog possessing a dansyl fluorophore attached to the C8 position of the adenine ring. This analog was shown to function as a cofactor for the adenine-N6-specific DNA methyltransferase MTaqI from *Thermus aquaticus*, resulting in the cofactor covalently attached to the exocyclic amino group (N6) of the target adenine located within sequence 5′-TCGA-3′ (Pljevaljčić et al. 2003). Subsequently, the groups of Rajsky and Comstock expanded the chemical scope of this approach by introducing 2-haloethyl *N*-mustard analogs, which are converted into aziridines in situ and thus are presumed to work by a similar mechanism (Weller and Rajski 2005, 2006; Townsend et al. 2009; Mai and Comstock 2011; Du et al. 2012; Ramadan et al. 2014). In the N-mustard cofactors, the N atom, which is equivalent to the sulfur atom of the sulfonium group in AdoMet, can in addition be used to attach a reactive chemical group (alkyne) (Weller and Rajski 2005) or a photocaging group (Townsend et al. 2009). Alternatively, retention of the amino acid moiety (present in AdoMet but absent in the N-adenosylaziridine analogs) renders enhanced cofactor-MTase affinity which gives a certain benefit of lower concentrations of cofactors that can be used in the labeling reactions (Weller and Rajski 2006; Du et al. 2012; Ramadan et al. 2014). Both the aziridine and N-mustard cofactors are obtained via multistep synthetic routes and can thus only be produced in specialized chemistry laboratories. Altogether, a variety of cofactor analogs have been produced containing reporter (biotin, fluorophores) or functional reactive groups (azide, alkyne) attached to the N6 or C8 or designed C7, positions of the adenine ring (see Table 1).

Fig. 1 Methyltransferase-directed sequence-specific labeling of DNA using synthetic analogs of the cofactor AdoMet. (*Left*) SMILing approach: covalent coupling of an aziridine (*upper*) or N-mustard (*lower*) cofactor carrying a functional or reporter group (*red sphere*) attached via a linker (*red line*) onto a target nucleobase (*blue*) in DNA. (*Right*) mTAG approach: transfer of a sulfonium-bound extended linear chain carrying an activating triple (*upper*) or double (*lower*) bond, a linker, and a functional or reporter group from a double-activated AdoMet analog onto a target nucleobase in DNA. N, random nucleotide; XXXXX, recognition sequence of the directing MTase

An important feature of the SMILing reaction is that the directing MTase remains tightly (although non-covalently) bound to the coupling product, which represents a chemically linked bisubstrate derivative entangling the enzyme. Therefore, stoichiometric amounts of an enzyme with respect to its target sites on target DNA are required for quantitative conversion, and additional steps may be necessary if the bound enzyme needs to be removed from the DNA. The attached cofactor represents a relatively bulky group, which may be a useful or inferior feature depending on the downstream applications.

The second DNA labeling approach is based on AdoMet analogs in which the sulfonium-bound methyl group of AdoMet is replaced with an extended side chain,

Table 1 Cofactor analogs for MTase-directed derivatization and labeling of DNA

Reactive or transferable moiety	Position of linker	Linker length	Functional or reporter group	Name	MTases used	Applications	References
SMILing							
N-Adenosyl-aziridines	Ade-N6	5	Biotin	6BAz	M.BseCI, M.TaqI	Positioning of nanoparticles on bacteriophage DNA Engineering synaptic junctions in DNA Optical mapping of DNA-binding proteins	Braun et al. (2008), Wilkinson et al. (2008), Kim et al. (2012)
		11	Cy3	6Cy3Az	M.TaqI	Transfection with fluorescently labeled plasmid DNA	Schmidt et al. (2008)
	Ade-C7	4	Biotin		M.HhaI, M.TaqI	Biotinylation and CpG methylation detection on plasmid DNA	Kunkel et al. (2015)
	Ade-C8	–	Azide		M.EcoRI, M.HhaI, M.SssI, M.TaqI	Derivatization and biotinylation of ODNs[a] MTase-directed DNA strand scission	Comstock and Rajski (2005a, b)
		5	Azide		M.EcoRI, M.TaqI	ODN biotinylation	Comstock and Rajski (2005a)
		5	Biotin		M.TaqI	Biotinylation of plasmid DNA	Pljevaljčić et al. (2004, 2007)
		6	Dansyl		M.TaqI	Fluorescent labeling of plasmid DNA	Pljevaljčić et al. (2003, 2004)

(continued)

Table 1 (continued)

Reactive or transferable moiety	Position of linker	Linker length	Functional or reporter group	Name	MTases used	Applications	References
N-Adenosyl-*N*-mustards	Mustard-N	1	Alkyne		M.EcoRI, M.TaqI	ODN derivatization	Weller and Rajski (2005)
	Ade-N6	1; 2; 4 3; 4	Alkyne Azide		M.HhaI, M.TaqI	Derivatization of plasmid DNA	Ramadan et al. (2014)
	Ade-C8	2 5	Alkyne Azide		M.HhaI, M.TaqI	Fluorescent labeling of plasmid DNA	Du et al. (2012)
		–	Azide		M.TaqI	Derivatization of plasmid DNA	Mai and Comstock (2011)

mTAG							
S-Propargyl analogs	S	1	Alkyne	2-Butynyl-SAM[b]	M.TaqI	DNA labeling and extraction	Artyukhin and Woo (2012)
		6	Alkyne	Ado-6-ethyne	eM.HhaI	Fluorescent labeling of plasmid DNA	Lukinavičius et al. (2013)
		6	Amine	Ado-6-amine	eM.HhaI, eM.SssI	Biotin labeling of DNA for epigenome profiling	Lukinavičius et al. (2013), Kriukienė et al. (2013)
		9	Amine	Ado-9-amine	eM.HhaI, M.TaqI	Fluorescent labeling of plasmid DNA	Lukinavičius et al. (2007)
		11	Amine	Ado-11-amine	eM2.Eco31I, eM.HhaI, eM.HpaII	Fluorescent labeling of plasmid and phage DNA Optical DNA mapping	Neely et al. (2010), Lukinavičius et al. (2012, 2013)
		6	Azide	Ado-6-azide	eM.HhaI, eM.SssI	Fluorescent labeling of plasmid DNA ex vivo Biotin labeling of DNA for epigenome profiling	Lukinavičius et al. (2013), Kriukienė et al. (2013)
		18	Biotin	Ado-18-biotin	eM.HhaI	Labeling of plasmid DNA	Urbanavičiūtė et al. UO[c], Plotnikova et al. (2014)
		20	TAMRA	AdoYnTAMRA	M.TaqI	Fluorescent labeling of phage DNA for optical strain typing	Grunwald et al. (2015)
S-Allyl analogs	S	3	Alkyne	AdoEnYn	M.FokI, M.TaqI, M.XbaI	Fluorescent labeling of phage DNA for optical DNA mapping	Vranken et al. (2014)

[a] *ODN* oligodeoxyribonucleotide
[b] This cofactor analog is also called AdoButyn in Tomkuvienė et al. (2012)
[c] *UO* unpublished observations

and only this part of the cofactor is transferred to the target nucleotide (Fig. 1). In AdoMet, the transferable methyl group is activated by the adjacent sulfonium center, and AdoHcy serves as the leaving group during the MTase-catalyzed S_N2 reaction. Replacement of the methyl group in AdoMet with larger aliphatic carbon chains had previously been attempted by Schlenk and Dainko (1975), who found that even short groups such as ethyl or propyl led to a drastic decline of transfer rates by MTases. A strongly decreased reaction rate observed with the saturated alkyl groups predominantly results from unfavorable steric effects within the transition state. In a joint effort, the Klimašauskas and Weinhold groups found that the efficiency of the reaction can be enhanced by placing π-orbitals near the reaction center (Dalhoff et al. 2006a). This activation was observed with synthetic AdoMet analogs carrying a double bond (allylic system) or a triple bond (propargylic system) next to the reactive carbon in the extended side chain (Fig. 1). Mechanistic considerations suggest that the π-orbitals in the unsaturated bond lower the energy barrier of the reaction via conjugative stabilization of a pentacoordinated S_N2 transition state. The discovery of the double-activated AdoMet analogs paved the way to a rapid development of a new approach termed methyltransferase-directed transfer of activated groups (mTAG).

Synthetic access to the mTAG cofactors appears somewhat easier as compared to the aziridine and N-mustard analogs, since they can be produced in a single step by chemical "recharging" of the cofactor product AdoHcy via regiospecific alkylation of its sulfur atom with a desired linear side chain. Suitable electrophilic side chains can sometimes be obtained directly from commercial sources, but certain cases may require advanced synthetic skill (Lukinavičius et al. 2007, 2013; Dalhoff et al. 2006b; Masevičius et al. 2016). Chemical synthesis typically yields the cofactor analogs as diastereomeric mixtures of *R*,*S*- and *S*,*S*-isomers, which can be chromatographically enriched in the enzymatically active *S*,*S*-isomer by reversed phase chromatography (Lukinavičius et al. 2013). Recently, a chemo-enzymatic synthesis of enantiomerically pure mTAG cofactors from corresponding methionine analogs and ATP using engineered methionine adenosyltransferases has been demonstrated (Singh et al. 2014), which can in principle be performed even in living cells (Wang et al. 2013).

Since only the extended sulfonium-bound side chain is transferred from the cofactor analog to DNA, these AdoMet analogs circumvent the problem of catalytic product release, which is unavoidable for the SMILing reactions. A variety of both allyl-based and propargyl-based analogs have been designed that carry unique chemical groups such as primary amine, alkyne, and azide or reporter groups (biotin, fluorophores) (see Table 1). Notably, although many MTases accept well both types of mTAG cofactors, some exhibit certain preferences with respect to the activating unsaturated bond (double or triple) or the side chain length. In particular, allylic cofactors have gained significant popularity with protein labeling (Peters et al. 2010; Islam et al. 2011, 2012; 2013; Wang et al. 2011; Blum et al. 2013; Wang et al. 2013; Bothwell and Luo 2014; Guo et al. 2014), whereas propargylic side chains are more preferably transferred by the C5-DNA MTases (Table 2). Unexpectedly, some of the propargyl cofactors containing an electronegative group

(amino or amido) at position four of the side chain were found to undergo a rapid loss of activity under physiological conditions. Further studies indicated that a close proximity of electron withdrawing groups makes the triple C-C bond highly susceptible to base-promoted addition of a water molecule. This problem was resolved by synthesis of a series of hex-2-ynyl cofactor analogs in which the separation between the electronegative group and the triple bond is increased from one to three carbon units (Lukinavičius et al. 2013). A similar mechanism has also been proposed for the fast inactivation of the AdoMet analog carrying a short unsubstituted prop-2-yn-1-yl side chain. In this case, the undesirable chemical reactivity of the triple bond has been diminished by synthetically replacing the sulfur atom in the highly electronegative sulfonium center with selenium (Bothwell et al. 2012; Willnow et al. 2012).

Both the SMILing and mTAG cofactors can be used for two-step or one-step labeling. A key advantage of the two-step labeling approach is the flexibility in manipulating the chemical parameters of the labeling reaction (linker length, conjugation chemistry, reporter group) by simply combining different cofactors and chemoselective reporter compounds. Alternatively, single-step labeling by direct attachment of a desired reporter group may be beneficial in situations when minimal sample manipulations, simplicity, and speed are required. However, besides this potential advantage, the one-step approach entails an added synthetic complexity to the cofactor analog, as reporter groups are typically larger and more complex than functional groups. Moreover, an increased steric bulk of the transferable side chain may also lead to a partial or complete impairment of the directing MTase (Table 2).

3 MTase Activity with the Synthetic Cofactor Analogs

Bacterial and archaeal DNA MTases generally exhibit a clearly defined sequence and base specificity. Bacterial type II DNA MTases (typically, single polypeptides of 250–400 residues) seem to be better suited for DNA labeling purposes as compared to the type I and III enzymes or mammalian DNA MTases, mostly due to their compact size and better enzymatic parameters (turnover rate, cofactor affinity, sequence fidelity, protein stability, etc.), although this general assumption does not preclude the existence of useful MTases derived from other than type II cohorts. Current listings of type II DNA MTases (REBASE, http://rebase.neb.com) count over 350 distinct recognition sequences ranging from 2 to 8 base pairs in length. Therefore, a wide repertoire of DNA sequences can potentially be targeted, which is in par with that of the widely used restriction endonucleases.

Naturally, DNA MTases have evolved for optimal performance with the natural cofactor AdoMet. The use of extended AdoMet analogs raises the question of steric limitations that may be imposed by the architectures of the active sites and cofactor-binding pockets of MTases. As mentioned above, the SMILing cofactors offer several potential anchoring points in the adenosine moiety (6, 7, and 8 positions of the adenine ring for the aziridines and additionally 5′-N for the N-mustards) that can be used for building a suitable extension carrying a desired functionality. This thus

Table 2 Activity of DNA methyltransferases with AdoMet analogs

Enzyme	Target sequence 5′– 3′	SMILing: Reactions performed[a]	SMILing: References	mTAG: Reactions performed[a]	mTAG: References
m6A-MTases					
M.TaqI wt	TCGA*	M F L	Pignot et al. (1998), Weller and Rajski (2005, 2006), Comstock and Rajski (2005a,b), Du et al. (2012), Mai and Comstock (2011), Ramadan et al. (2014), Pljevaljčić et al. (2003, 2004, 2007), Braun et al. (2008), Schmidt et al. (2008), Wilkinson et al. (2008), Kunkel et al. (2015)	M F L	Dalhoff et al. (2006a), Lukinavičius et al. (2007), Vranken et al. (2014), Artyukhin and Woo (2012), Grunwald et al. (2015)
M.BseCI	ATCGAT	L	Braun et al. (2008), Wilkinson et al. (2008), Kim et al. (2012)		
M.EcoRI	GAATTC	M F	Weller and Rajski (2005, 2006), Comstock and Rajski (2005a)	N	Vranken et al. (2014)
M.FokI	GGATG/ CATCC			F	Vranken et al. (2014)
M.XbaI	TCTAGA			F	Vranken et al. (2014)
M.EcoDam	GATC			N	Vranken et al. (2014)
M.PstI	CTGCAG			N	Vranken et al. (2014)
m5C-MTases					
M.HhaI wt	GCGC	M F L	Weller and Rajski (2006), Pljevaljčić et al. (2004), Comstock and Rajski (2005a,b), Ramadan et al. (2014), Du et al. (2012), Kunkel et al. (2015)	M	Dalhoff et al. (2006a)
Q82A/N304A				F	Lukinavičius et al. (2007, 2012, 2013)
Y254S/N304A				F	Lukinavičius et al. (2012)

Q82A/Y254S/N304A				F L	Neely et al. (2010), Lukinavičius et al. (2012, 2013), Urbanavičiūtė et al. (unpublished)
M.HhaI ΔL2–14	GCG			M F	Gerasimaitė et al. (2009), Gerasimaitė et al. (unpublished)
Q82A/N304A				L	Urbanavičiūtė et al. (unpublished)
M.SssI wt	CG	M F	Weller and Rajski (2006), Comstock and Rajski (2005b)	N	Vranken et al. (2014)
Q142A/N370A				F	Kriukienė et al. (2013)
M.HpaII wt	CCGG	M	Comstock and Rajski (2005a)	M	Lukinavičius et al. (2012)
Q1042A/N335A				F	Lukinavičius et al. (2012)
M2.Eco31I wt	GGTCTC			F	Lukinavičius et al. (2012)
N127A/Q233A				F	Lukinavičius et al. (2012)
M.BsaHI wt	GRCGYC			N	Vranken et al. (2014)
Q75A/N274A/V220S				N	Vranken et al. (2014)
4mC-MTases					
M.BcnIB	CCSGG	M	Pljevaljčić et al. (2004)	M	Dalhoff et al. (2006a)
M.BamHI	GGATCC	M	Du et al. (2012)		
M.PvuII	CAGCTG			N	Vranken et al. (2014)

[a]*N* none or low alkyltransferase activity, *M* modification by transfer of a core unit (SMILing) or a short nonfunctional moiety (mTAG), *F* derivatization with a functional group (two-step labeling possible), *L* labeling with a reporter group in one step, * the target nucleotide is underlined

offers several chemical options for designing suitable cofactor for particular MTases. In the mTAG cofactors, there is only one attachment point, and the chemical variability of the side chain is basically limited to either the allyl or propargyl moieties, which demand quite distinct geometries of the cofactor pocket in both the ground and transition states. Of course, the length of the side chains and other chemical features can also influence the reaction to some extent, but such effects decline with increasing distance from the active site.

Representatives of all three classes of bacterial DNA MTases (m6A, m4C, and m5C forming enzymes) showed activity with certain types of extended cofactor analogs (see Table 2). On one end of the spectrum is M.TaqI, which demonstrated high tolerance with respect to a wide range of SMILing and mTAG cofactors examined. More typically though, the efficiency of mTAG transalkylations with wild-type enzymes is insufficient for routine applications. For C5-MTases, this issue was approached by engineering of the cofactor pocket of a well-characterized representative of the class, M.HhaI (Lukinavičius et al. 2012). The engineering effort was guided by a structure-based model of an M.HhaI-DNA-butynyl cofactor complex (Fig. 2), which suggested that the side chains of residues Gln82 and Asn304 (located in conserved sequence motives IV and X) and Tyr254 (located in the so-called variable region) might sterically interfere with the extended transferable side chain, precluding cofactor binding or its proper orientation for catalysis. These three positions were therefore selected for steric engineering (Ala or Ser replacements). It turned out that double and triple replacements conferred substantial improvements of the transalkylation activity and a reduction of the methyltransferase activity in M.HhaI. The achieved turnover rates permit complete derivatization of DNA in 15–30 min, which makes the reaction suitable for routine laboratory applications. Detailed studies of the mutants showed that these replacements substantially enhance the rate of alkyl transfer and also reduce the enzyme affinity toward the natural cofactor AdoMet and

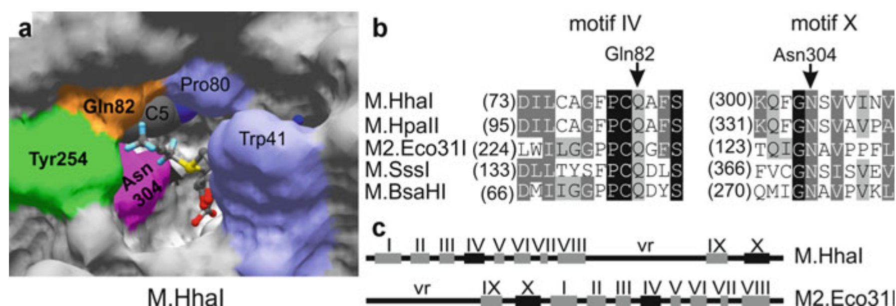


Fig. 2 Structure-guided engineering of DNA cytosine-C5 methyltransferases for the mTAG transalkylation reactions. (**a**) Model of an extended propargylic cofactor analog (AdoButyn, shown in ball and stick) bound in the active site of the HhaI MTase (based on M.HhaI-DNA-AdoMet ternary complex X-ray structure, PDB code 6mht, shown as space fill). An *arrow points* at the transferable carbon atom. (**b**) Sequence alignment of regions corresponding to IV and X conserved motifs of sterically engineered prokaryotic cytosine-C5 MTases. *Arrows* indicate positions corresponding to Gln82 and Asn304 of M.HhaI. (**c**) Permutation of conserved motifs and the variable region (vr) in M.HhaI, M.HpaII, M.SssI, M.BsaHI (*top*), and M2.Eco31I (*lower*) DNA methyltransferases (Adapted from Lukinavičius et al. (2012))

its product AdoHcy. The catalytic transfer of butyn-2-yl and pentyn-2-yl groups by the triple mutant was faster than its methyl transfer activity, indicating that the engineered enzyme (eM.HhaI) was turned into an alkyltransferase. Importantly, eM.HhaI can efficiently utilize extended synthetic analogs even in the presence of AdoMet (which is naturally abundant in cells and cell lysates), opening new ways for targeted covalent deposition of reporter groups onto DNA for a variety of ex vivo and in vivo applications (Lukinavičius et al. 2013). In line with these findings, a substantial improvement of the transalkylation activity was also observed in another engineered version of M.HhaI, which was designed to target nonsymmetrical GCG sites. Directed evolution of the MTase aimed at enforcing the new sequence specificity resulted in the above-described Tyr254Ser mutation and additional deletions in the vicinity of the cofactor-binding pocket (Gerasimaitė et al. 2009).

The high structural conservation of C5-MTases suggested that other orthologs can be similarly engineered based on sequence alignment even in the absence of crystal structures. Indeed, the double alanine mutants involving conserved motifs IV and X led to a significant improvement of the transalkylation activity with a wide range of propargyl-based cofactor analogs by M2.Eco31I and M.HpaII, which recognize hexanucleotide and tetranucleotide target sites, respectively (Lukinavičius et al. 2012), as well as by M.SssI acting on the 5′-CG-3′ dinucleotide (Kriukienė et al. 2013). On the other hand, analogous replacements in M.BsaHI showed no significant improvement in the transfer of allyl-based extended groups onto DNA (Vranken et al. 2014). This appears to agree with the observed weaker acceptance of double-bond cofactors by the engineered M.HhaI variants (Lukinavičius, Lapinaitė, Klimašauskas, unpublished observations), suggesting that the triple-bond cofactors are generally better compatible with the C5-MTases. For more details on MTase design, see the chapter by Laurino and Tawfik in this book.

4 Implementation of MTase-Directed Labeling of DNA

Sequence-specific covalent derivatization and labeling of DNA has potentially opened new avenues in DNA research, diagnostics, and bionanotechnology. However, along with methodological developments of the MTase-directed labeling reactions, the properties and practical value of such covalently modified DNA that suddenly became available needed to be assessed. Many experimental demonstrations involving various covalently tethered reporter and reactive groups have been performed at the level of oligonucleotides, PCR fragments, and then plasmid DNA. These studies can be grouped into those that exploited covalent derivatizations for (1) general covalent labeling of DNA or (2) analysis of particular DNA sites or sequences.

4.1 General Covalent Labeling of DNA

Soon after convincing demonstrations that both SMILing and mTAG techniques can achieve high sequence specificity of label incorporation into plasmid DNA (Pljevaljčić et al. 2007; Lukinavičius et al. 2007), the behavior of covalently labeled

plasmid DNA was examined in transfected cells. For example, an aziridine-based cofactor with a Cy3 fluorophore was used for labeling of pUC19 and pBR322 plasmids with M.TaqI; the plasmids were successfully transfected and optically tracked in mammalian cells (Schmidt et al. 2008). Independently, mTAG-derivatized plasmids were shown to have transformation efficiencies similar to unmodified plasmid controls in *E. coli* cells (Lukinavičius et al. 2012). Moreover, sequence-specific mTAG click-labeling of endogenous plasmid DNA using eM.HhaI and Ado-6-azide cofactor followed by strain-promoted azide-alkyne cycloaddition (SPAAC) of a cyclooctyne probe was demonstrated in bacterial cell extracts (Lukinavičius et al. 2013). Altogether these experiments demonstrated a high biological tolerance (bioorthogonality) of both types of covalent modifications pointing at potential suitability of this approach for in vivo studies. The exceptional selectivity of DNA MTases toward DNA can also be used for covalent capture and extraction of DNA from complex mixtures (Artyukhin and Woo 2012). Modification of DNA with alkynyl groups using mTAG technique and further covalent immobilization through copper (I)-catalyzed azide-alkyne cycloaddition (CuAAC) reaction on azide-coated surfaces permits its further manipulations and compatibility with downstream reactions. The approach of DNA separation from other biomolecules, including RNA, showed sensitivity and selectivity unprecedented in other DNA extraction methods.

Altogether, for the purpose of general DNA labeling, the MTase-directed methods offer important advantages over random chemical labeling or other commonly used methods:

- Elimination of uncertainties related to loss of non-covalently bound labels in cells or in vitro experiments
- Control of the labeling density and positioning of reporter groups around (or away from) functional sites by selecting appropriate MTases
- High flexibility in selecting functional and reporter groups
- Covalent integrity of DNA strands (preserved supercoiling of plasmid DNA)
- Biological orthogonality of the underlying modifications

4.2 DNA Labeling for Analysis of Particular DNA Sites or Sequences

Another layer of utility of the MTase-directed labeling is related to the exploitation of individual labeled sites in DNA. One such area is the construction of DNA-based nanostructures. Braun et al. (2008) used biotinylated aziridine cofactors together with M.TaqI and M.BseCI (recognizing tetranucleotide and hexanucleotide sequences, respectively) for biotin labeling and subsequent targeted deposition of gold nanoparticles on model kilobase-sized DNA fragments via biotin-streptavidin interaction. Wilkinson et al. (Wilkinson et al. 2008) used similar tools to engineer

synaptic three- and four-way junctions in PCR-derived DNA, which were unequivocally visualized using single-molecule AFM imaging. These examples demonstrated the capacity of MTase-directed labeling for controlled manipulation of nanoparticles on DNA scaffolds and directing the bottom-up assembly of nanomaterials, which await their further technological implementation in many fields related to molecular electronics, biosensors, optical waveguides, etc.

4.2.1 Optical Mapping of DNA Sequences

AFM (Wilkinson et al. 2008) or EM (Kunkel et al. 2015) visualization of several bulky nano-objects along the irregular contour of a DNA molecule spotted on a mica surface gives a nice qualitative illustration, but is poorly suited for fast parallel analysis of DNA molecules containing a large number of target sites. Direct determination of physical distances (positioning) between the specific sites becomes possible on stretched-out DNA molecules, leading to a visual pattern characteristic of that particular DNA. Such a linear representation of a DNA sequence, called optical map, can be read as a barcode and analyzed with a high degree of automation (Fig. 3a). Direct single-molecule analysis of large DNA fragments, which far exceed the read length of widely used sequencing technologies, provides valuable genomic information for the identification of structural or copy number variations and assists with DNA sequence assembly or rapid strain typing (reviewed in Levy-Sakin and Ebenstein (2013)). Several partially or fully automated optical DNA mapping platforms (BioNano Genomics, GenomicVision, PathoGenetix, OpGen) are already available. However, implementation of various known methods for optical DNA map generation is dependent on many technical parameters related to the degree and accuracy of label incorporation, repertoire of available target sites, covalent continuity of labeled DNA strands, inhomogeneous stretching, chemical and physical stability of the fluorophores, resolution and speed of signal readout, etc. Existing methods for specific visual pattern generation in optical DNA mapping include restriction map generation (Teague et al. 2010), nick translation (Lam et al. 2012), and probe hybridization (Weier et al. 1995), all of which suffer from one or more of the above-listed limitations (discussed in Zohar and Muller (2011) and Levy-Sakin and Ebenstein (2013)).

The MTase-based approaches, owing to their unique combination of high specificity, covalent bonding, and DNA strand preservation, appear particularly suited for this purpose. In a proof of principle study, two-step mTAG labeling was employed to attach fluorophores on 215 HhaI sites in bacteriophage lambda DNA (48.5 kb, see Fig. 3a) (Neely et al. 2010). The labeling employed engineered M.HhaI and a cofactor bearing a transferable linear side chain with a terminal amino group followed by a chemoselective attachment of an Atto647N dye. The DNA molecules were stretched by combing onto polymer-coated coverslips using an evaporating droplet technique. Positions of fluorophores along individual DNA strands were recorded at sub-diffraction resolution (10 nm or just 20 bp) using dSTORM imaging, which utilized photobleaching of the fluorophores to ensure that single emitters

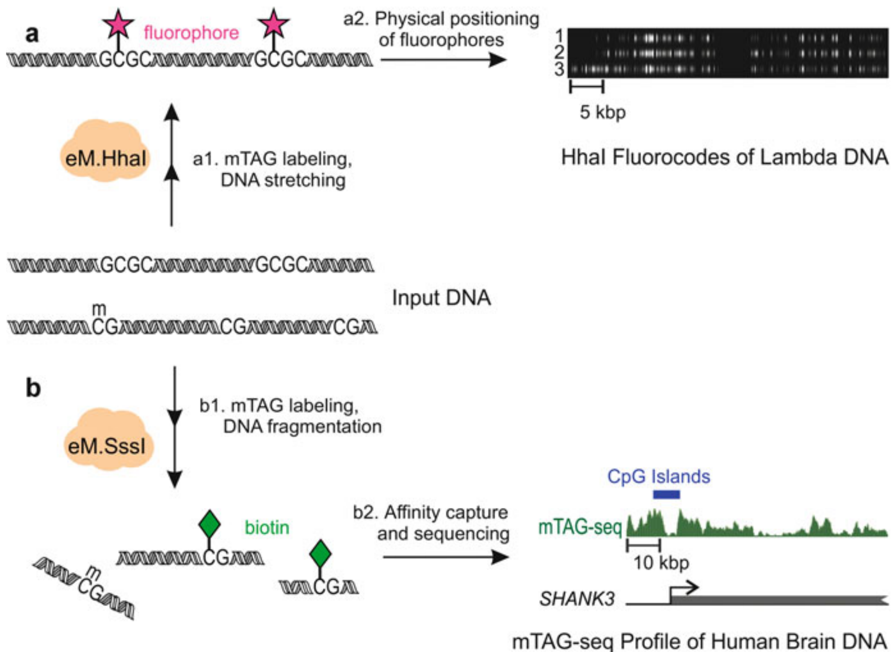


Fig. 3 Major applications of methyltransferase-directed labeling in genome studies. (**a**) Optical DNA mapping using fluorescent mTAG labeling (Neely et al. 2010). A two-step mTAG reaction involving an engineered version of HhaI methyltransferase (eM.HhaI) was used for fluorescent labeling of GCGC sites in bacteriophage lambda DNA (*a1*). The labeled DNA molecules were stretched out by combing, and positions of the fluorophores on individual DNA molecules were determined using super-resolution imaging (*a2*). Illustration in the upper right corner shows 1 and 2 (experimental consensus fluorocodes derived by using different processing parameters) and 3 (in silico generated (theoretical) reference map) (Adapted from Neely et al. 2010). (**b**) DNA "unmethylome" profiling by covalent mTAG labeling of unmodified CG sites (Kriukienė et al. 2013). Unmodified CG sites in fragmented genomic DNA from the human brain (50–300 bp fragments) were biotin tagged in a two-step mTAG labeling reaction involving an engineered variant of the SssI methyltransferase (eM.SssI) (*b1*). Biotin-tagged fragments were affinity enriched and sequenced to produce a genome-wide profile of unmodified CG sites (*b2*). Illustration in the lower right corner shows a genome browser view of mTAG-seq data over a part of *SHANK3* gene including its promoter, overlapping with a predicted CpG island region (Adapted from Kriukienė et al. 2013). C^m refers to naturally modified cytosine (mC/hmC/fC/caC)

are isolated and their positions accurately determined. While an average density of localized sites of approximately one per 650 bases represented a significant advance compared to other DNA mapping technologies, this density was achieved by assigning only 34 % of the 215 available HhaI sites on the DNA. To further increase the number of fluorophores in the experimentally derived map, a consensus "fluorocode" encompassing nearly 90 % of the sites (density 1/270 bp) was generated from 20 automatically aligned molecules.

To further explore the technicalities of the labeling reaction for improved optical mapping of DNA, a cofactor carrying a short allylic side chain with a terminal

alkyne group (AdoEnYn) was used along with CuAAC-mediated attachment of a fluorophore (Vranken et al. 2014). Eleven MTases were screened for activity with this cofactor, of which three adenine-specific enzymes were found to be active. The CuAAC-based approach generated bacteriophage T7 fluorocodes with labeling efficiency reaching 70 %; however, the authors noted substantial degradation of DNA in the presence of Cu (I), which precluded generation of full-length labeled DNA molecules. Given that a wide selection of mTAG cofactors along with a proven set of engineered and wt MTases is now available, a successful implementation of this approach (e.g., via a copper-free AAC reaction or one-step labeling) seems just around the corner.

To this end, one-step mTAG labeling with M.TaqI and a propargylic cofactor carrying a linker-bound TAMRA fluorophore was used to label over 200 target sites on lambda DNA (Grunwald et al. 2015) followed by physical stretching of the DNA molecules in commercial microchip-based nanochannel arrays. A fluorescent signature of DNA was generated by "conventional resolution" imaging, i.e., measuring the amplitude modulations of fluorescence intensity along its length rather than isolated fluorescent spots. This resulted in lower resolution density profiles characteristic of an underlying DNA molecule. Although nearly quantitative labeling was presumably achieved, images of individual molecules did not appear identical, pointing to the stochastic nature of single-molecule measurements. Nevertheless, the generated consensus profiles permitted a clear distinction between the two types of bacteriophage DNA used. The generality, rapidness, and high throughput make this concept promising for routine applications in strain typing assays.

The SMILing technique proved instrumental for sparse fluorescent labeling of the T7 bacteriophage genome (three sites per ~40 kb) using M.BseCI and a biotinylated aziridine-based cofactor in a two-color experiment. Streptavidin-coated quantum dots were attached at the biotinylated target sites as genomic reference tags to aid mapping the locations of non-covalently bound RNA polymerase molecules labeled with differently colored probes (Kim et al. 2012). Introduction of such reference tags allowed a higher precision in the assignment of the RNAP-binding sites relative to localization based on their distance from DNA ends.

Altogether, the above examples demonstrate that MTase-directed approaches provide a valuable addition to the toolbox of sequence-specific labeling techniques, which will accelerate the development of automated high-throughput technologies for optical DNA mapping.

4.2.2 Applications of MTase-Directed Labeling in Epigenomics

Yet another emerging direction of practical utility for MTase-directed sequence-specific labeling is the analysis of modified target sites in natural DNA. As discussed in the previous chapters, the prevalent covalent modification of the genome is sequence-specific methylation of cytosine and adenine residues. In higher eukaryotes including mammals, DNA cytosine-5 methylation predominantly occurs at CpG dinucleotides and acts as a stably inherited modification affecting gene regulation and cellular differentiation. Aberrant DNA methylation is an early and fundamental event in pathogenesis of many human diseases, including cancer (reviewed

in Jones (2012)). Besides 5mC, other DNA modifications have been discovered recently, giving rise to extensive discussions of their potential roles as epigenetic marks (reviewed in Kriukienė et al. (2012)). To gain mechanistic insights into the dynamics and function of DNA methylation, genome-wide analyses of DNA modification patterns have been performed in different organisms and cell types employing a variety of profiling techniques (Weber et al. 2005; Schumacher et al. 2006; Bock 2008; Harris et al. 2010). During the past few years, chemical tagging of modification sites has been adapted for in vitro epigenome studies. Covalent derivatization of modified residues permitted incorporation of reactive azide, keto, or primary amine groups followed by chemoselective conjugation of biotin (Song et al. 2011; Pastor et al. 2011; Zhang et al. 2013).

The key concept of using MTase-directed labeling for analysis of mammalian genomic DNA lies in selective covalent tagging of the unmodified fraction of CG sites, whereas the naturally modified sites will remain untagged due to preexisting modification of the target residue. As unmodified cytosines represent a smaller proportion of CG sites compared with methylated ones (depending on the tissue, 65–80 % of cytosines in the human genome are methylated (reviewed in Suzuki and Bird (2008)), analysis of this smaller, unmethylated DNA fraction may reduce the number of statistical comparisons and is more sensitive for detecting subtle epigenetic changes. An early attempt to analyze DNA methylation sites through targeted DNA scission (Comstock and Rajski 2005b) used derivatization of model oligodeoxynucleotide substrates with M.TaqI or M.HhaI and an azide-bearing aziridine cofactor, which was further subjected to the Staudinger ligation with triarylphosphines derivatized with phenanthroline. Presentation of these duplexes with Cu (II) promoted strand scission at the vicinity of the base modified by the enzyme. However, this chemistry leads to extensive DNA damage, and the remaining DNA fragments are not readily analyzed by modern sequencing techniques.

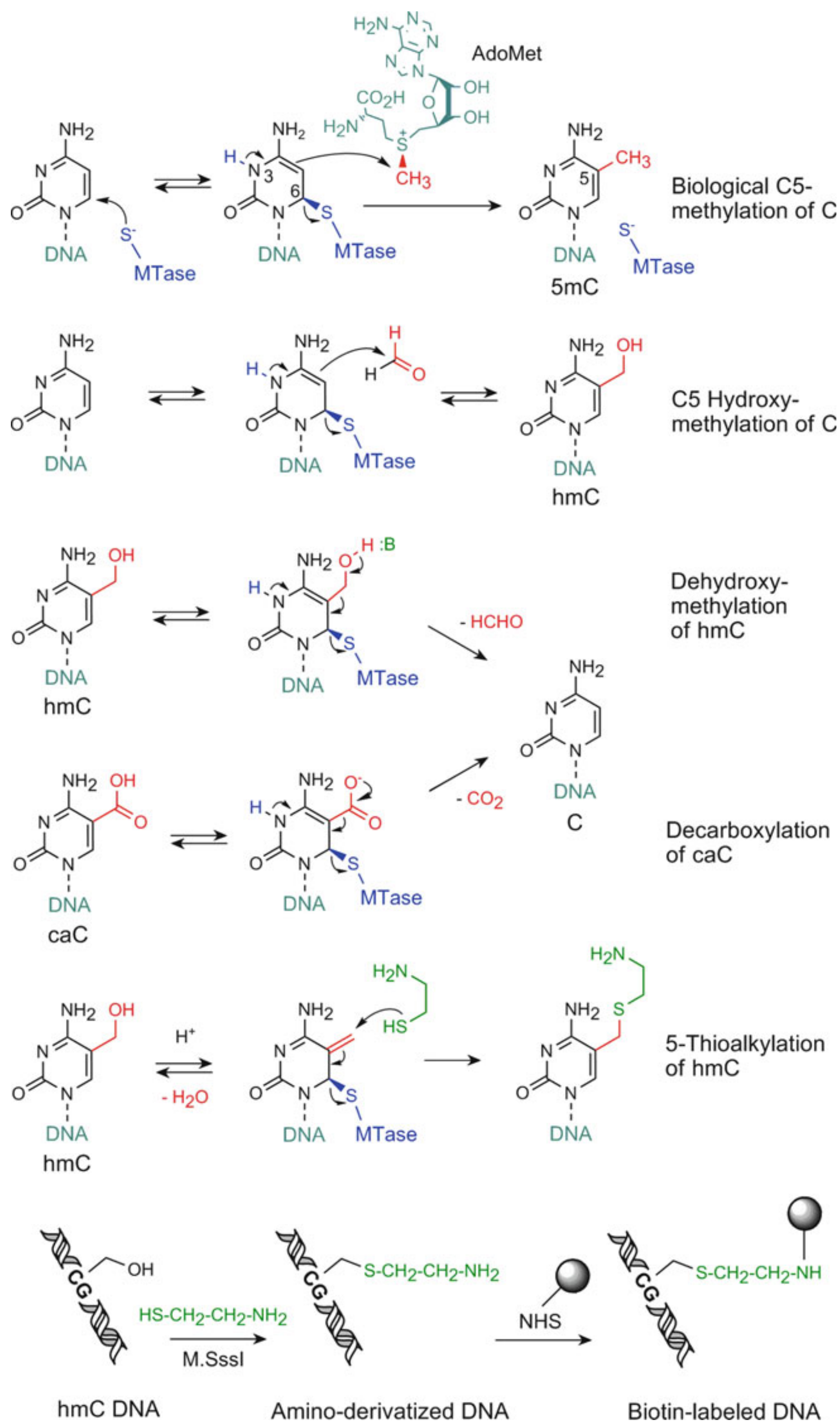
A more recent demonstration of chemo-enzymatic profiling of the unmodified fraction of the genome (named “unmethylome”) was based on selective covalent capture of CG sites (Kriukienė et al. 2013) (Fig. 3b). Covalent tagging of DNA was performed using the engineered version of M.SssI (see Sect. 3) and a synthetic mTAG type analog of AdoMet cofactor carrying a terminal amine or azide group. In the next step, conventional chemoselective coupling of the amine group with an NHS-biotin probe or, alternatively, SPAAC of the attached azide group with a dibenzocyclooctyne biotin reagent was employed. Biotin-labeled DNA fragments were then enriched on streptavidin beads and analyzed on tiling DNA microarrays (mTAG-chip) or by next-generation sequencing (mTAG-seq). Pilot profiling studies of human DNA samples from cultured cells and tissues demonstrated that this approach offers nanogram sensitivity and permits identification of unmethylated CG sites genome wide with high precision and reproducibility. Moreover, mTAG-seq can be considered not only as a powerful and economical alternative but also as a complementary technique to 5mC-specific methods such as affinity-based method MeDIP (Weber et al. 2005) and, likely, to TAmC-seq, a method for covalent

derivatization and analysis of methylated cytosines (Zhang et al. 2013). Most recent modification of the approach includes covalent tethering of a priming oligonucleotide to the tagged nucleotides, which offers a particularly cost-effective direct genomic mapping of each unmodified CG site at near single-base resolution (Kriukienė et al., manuscript in preparation).

5 Cofactor-Independent MTase-Directed Labeling

In addition to their well-characterized catalytic activity, DNA C5-MTases were found to catalyze atypical reactions involving non-cofactor-like substrates. As mentioned above, the C5-MTases use a covalent mechanism for nucleophilic activation of their target cytosine residues. The transiently generated activated cytosine intermediate is not only active toward AdoMet or its synthetic analogs but can also attack other exogenous electrophiles such as aliphatic aldehydes, yielding corresponding 5-α-hydroxyalkylcytosines (Liutkevičiūtė et al. 2009) (Fig. 4a). The reactions occur under fairly mild conditions and retain the high sequence and base specificity characteristic of bacterial DNA MTases. The coupling with formaldehyde yields 5-hydroxymethylcytosine (hmC), which is a naturally occurring cytosine modification in mammalian DNA (Kriaucionis and Heintz 2009; Tahiliani et al. 2009). Although the hydroxymethyl groups themselves are not good chemical reporters, they add a unique functionality to DNA (analogous to benzylic hydroxyl) that can be exploited for chemical or enzymatic derivatization. For example, a mild oxidation to formyl or keto groups would enable a further conjugation with compounds carrying hydrazine or hydroxylamine functions (Prescher and Bertozzi 2005). Alternatively, hmC residues can be enzymatically glucosylated (Gommers-Ampt and Borst 1995), thereby permitting selective DNA labeling through application of glycan modification/recognition techniques (Chittaboina et al. 2005; Song et al. 2011).

Curiously, it was also found that the covalent activation of 5-substituted cytosine residues present at the target position of a C5-MTase can lead to their conversion into unmodified cytosine (Fig. 4a). This reaction does occur with hmC and 5-carboxylcytosine, but was not observed with 5-formylcytosine (Liutkevičiūtė et al. 2009, 2014). The MTase-activated hmC in DNA can also undergo condensation with exogenous aliphatic thiols and selenols yielding corresponding 5-alkylchalcogenomethyl derivatives (Liutkevičiūtė et al. 2011) (Fig. 4a). Since this MTase-directed derivatization reaction is not possible at 5-methylated and unmodified cytosine residues, it appears well suited for selective covalent capture of 5-hydroxymethylated-CG sites in mammalian genomic DNA. As a proof of concept, C5-MTase-directed derivatization of hmC with cysteamine and subsequent amine-selective biotin labeling (Fig. 4b) was demonstrated on plasmid DNA and model DNA fragments (Liutkevičiūtė et al. 2011) and was subsequently implemented in a commercial analytical tool (EpiJet 5-hmC Enrichment Kit). Moreover, M.HhaI and M.SssI have recently been shown to render sequence-specific



AdoMet
Biological C5-methylation of C
5mC
C5 Hydroxy-methylation of C
hmC
Dehydroxy-methylation of hmC
- HCHO
- CO2
C
Decarboxylation of caC
caC
5-Thioalkylation of hmC
H+
- H2O
HS-CH2-CH2-NH2
M.SssI
NHS
S-CH2-CH2-NH2
S-CH2-CH2-NH
hmC DNA
Amino-derivatized DNA
Biotin-labeled DNA

conjugation of short Cys-containing peptides to hmC-containing DNA (Serva and Lagunavičius 2015).

Altogether, the presented variety of atypical reactions demonstrate a high catalytic plasticity of DNA C5-MTases and offer additional ways for sequence-specific derivatization of canonical and modified bases within DNA. As compared to the cofactor-based reactions, these reactions typically require simpler and less expensive compounds, thereby avoiding multistep syntheses of AdoMet analogs.

6 Conclusions and Outlook

MTase-directed labeling of DNA is an enabling technology with many unique demonstrated applications. Due to its relative simplicity, robustness, and wide-range applicability, this approach is becoming a method of choice where targeted covalent derivatization of DNA is required. Although certain technical questions still require attention, the rapidly growing popularity indicates that the field is approaching its maturity stage. The two most developed applications of the method are optical DNA mapping and analysis of epigenetic states in mammalian DNA; both methods are now entering the phases of automation and commercial exploitation, and no doubt will soon become commonly used technologies. Another important area of research that is poised to see a rapid bloom in the near future is DNA labeling in biological systems and in living cells. Currently, two main obstacles can be envisioned: (1) entrance/delivery of cofactor analogs into cells and (2) design of highly orthogonal cofactor-MTases pairs for allele-specific labeling. The first issue can be addressed by harnessing cell delivery systems, which are widely used to cargo a variety of other molecules across the cell membrane (Janib et al. 2010; Falanga et al. 2015), or by enzymatic production of cofactor analogs in situ from corresponding methionine analogs, which show superior wall penetration properties (Wang et al. 2013; Singh et al. 2014).

Acknowledgments This work was supported by grants from the National Institutes of Health (HG007200) and the Research Council of Lithuania (MIP-45/2013).

Fig. 4 Cofactor-independent methyltransferase-directed sequence-specific derivatization of DNA. (**a**) Transformations of a target cytosine catalyzed by DNA C5-MTases. Biological methylation by C5-MTases occur via an S_N2 reaction between an activated cytosine intermediate (ACI) and cofactor AdoMet, yielding 5-methylcytosine (5mC) (biological C5-methylation). The ACI can undergo nucleophilic addition reaction with short aliphatic exogenous aldehydes, which in the case of formaldehyde yields hmC (C5-hydroxymethylation). In the reverse reaction, hmC residues can be converted to unmodified cytosines in DNA by the enzyme (dehydroxymethylation of hmC). Similarly, 5-carboxycytosine (caC) can be converted to cytosine (decarboxylation of caC). hmC residues, including those naturally occurring in DNA, can undergo further methyltransferase-directed condensation with thiol or selenol reagents to give stable 5-alkyl chalcogenomethyl derivatives (5-thioalkylation of hmC). Modifying reagents are shown in *red* and *green* (thiol), and C5-MTase and its catalytic moieties are shown in *blue*. (**b**) MTase-directed covalent amino derivatization and labeling of hmCG dinucleotides in DNA with biotin (shown as a ball)

References

Artyukhin AB, Woo YH. DNA extraction method with improved efficiency and specificity using DNA methyltransferase and “click” chemistry. Anal Biochem. 2012;425(2):169–74.

Blum G, Islam K, Luo M. Bioorthogonal profiling of protein methylation (BPPM) using an azido analog of S-adenosyl-L-methionine. Curr Protoc Chem Biol. 2013;5(1):45–66.

Bock C, Walter J, Paulsen M, Lengauer T. Inter-individual variation of DNA methylation and its implications for large-scale epigenome mapping. Nucleic Acids Res. 2008;36(10):e55.

Bothwell IR, Islam K, Chen Y, Zheng W, Blum G, Deng H, et al. Se-adenosyl-L-selenomethionine cofactor analogue as a reporter of protein methylation. J Am Chem Soc. 2012;134(36): 14905–12.

Bothwell IR, Luo M. Large-scale, protection-free synthesis of Se-adenosyl-L-selenomethionine analogues and their application as cofactor surrogates of methyltransferases. Org Lett. 2014;16(11):3056–9.

Braun G, Diechtierow M, Wilkinson S, Schmidt F, Huben M, Weinhold E, et al. Enzyme-directed positioning of nanoparticles on large DNA templates. Bioconjug Chem. 2008;19(2):476–9.

Chittaboina S, Xie F, Wang Q. One-pot synthesis of triazole-linked glycoconjugates. Tetrahedron Lett. 2005;46(13):2331–6.

Comstock LR, Rajski SR. Conversion of DNA methyltransferases into azidonucleosidyl transferases via synthetic cofactors. Nucleic Acids Res. 2005a;33(5):1644–52.

Comstock LR, Rajski SR. Methyltransferase-directed DNA strand scission. J Am Chem Soc. 2005b;127(41):14136–7.

Dalhoff C, Lukinavičius G, Klimašauskas S, Weinhold E. Direct transfer of extended groups from synthetic cofactors by DNA methyltransferases. Nat Chem Biol. 2006a;2(1):31–2.

Dalhoff C, Lukinavičius G, Klimašauskas S, Weinhold E. Synthesis of S-adenosyl-L-methionine analogs and their use for sequence-specific transalkylation of DNA by methyltransferases. Nat Protoc. 2006b;1(4):1879–86.

Du Y, Hendrick CE, Frye KS, Comstock LR. Fluorescent DNA labeling by N-mustard analogues of S-adenosyl-L-methionine. Chembiochem. 2012;13(15):2225–33.

Falanga A, Galdiero M, Galdiero S. Membranotropic cell penetrating peptides: the outstanding journey. Int J Mol Sci. 2015;16(10):25323–37.

Gerasimaitė R, Vilkaitis G, Klimašauskas S. A directed evolution design of a GCG-specific DNA hemimethylase. Nucleic Acids Res. 2009;37(21):7332–41.

Gommers-Ampt JH, Borst P. Hypermodified bases in DNA. FASEB J. 1995;9(11):1034–42.

Grunwald A, Dahan M, Giesbertz A, Nilsson A, Nyberg LK, Weinhold E, et al. Bacteriophage strain typing by rapid single molecule analysis. Nucleic Acids Res. 2015;43(18):e117.

Guo H, Wang R, Zheng W, Chen Y, Blum G, Deng H, et al. Profiling substrates of protein arginine N-methyltransferase 3 with S-adenosyl-L-methionine analogues. ACS Chem Biol. 2014; 9(2):476–84.

Harris RA, Wang T, Coarfa C, Nagarajan RP, Hong C, Downey SL, et al. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. Nat Biotechnol. 2010;28(10):1097–105.

Hymbaugh Bergman SJ, Comstock LR. N-mustard analogs of S-adenosyl-L-methionine as biochemical probes of protein arginine methylation. Bioorg Med Chem. 2015;23(15):5050–5.

Holstein JM, Schulz D, Rentmeister A. Bioorthogonal site-specific labeling of the 5′-cap structure in eukaryotic mRNAs. Chem Commun (Camb). 2014;50(34):4478–81.

Islam K, Bothwell I, Chen Y, Sengelaub C, Wang R, Deng H, et al. Bioorthogonal profiling of protein methylation using azido derivative of S-adenosyl-L-methionine. J Am Chem Soc. 2012;134(13):5909–15.

Islam K, Chen Y, Wu H, Bothwell IR, Blum GJ, Zeng H, et al. Defining efficient enzyme-cofactor pairs for bioorthogonal profiling of protein methylation. Proc Natl Acad Sci U S A. 2013;110(42):16778–83.

Islam K, Zheng W, Yu H, Deng H, Luo M. Expanding cofactor repertoire of protein lysine methyltransferase for substrate labeling. ACS Chem Biol. 2011;6(7):679–84.

Janib SM, Moses AS, MacKay JA. Imaging and drug delivery using theranostic nanoparticles. Adv Drug Deliv Rev. 2010;62(11):1052–63.

Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012;13(7):484–92.

Kim S, Gottfried A, Lin RR, Dertinger T, Kim AS, Chung S, et al. Enzymatically incorporated genomic tags for optical mapping of DNA-binding proteins. Angew Chem Int Ed Engl. 2012;51(15):3578–81.

Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science. 2009;324(5929):929–30.

Kriukienė E, Liutkevičiūtė Z, Klimašauskas S. 5-Hydroxymethylcytosine--the elusive epigenetic mark in mammalian DNA. Chem Soc Rev. 2012;41(21):6916–30.

Kriukienė E, Labrie V, Khare T, Urbanavičiūtė G, Lapinaitė A, Koncevičius K, et al. DNA unmethylome profiling by covalent capture of CpG sites. Nat Commun. 2013;4:2190.

Kunkel F, Lurz R, Weinhold E. A 7-deazaadenosylaziridine cofactor for sequence-specific labeling of DNA by the DNA cytosine-C5 methyltransferase M. HhaI. Molecules. 2015;20(11): 20805–22.

Lam ET, Hastie A, Lin C, Ehrlich D, Das SK, Austin MD, et al. Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nat Biotechnol. 2012;30(8): 771–6.

Lee BW, Sun HG, Zang T, Kim BJ, Alfaro JF, Zhou ZS. Enzyme-catalyzed transfer of a ketone group from an S-adenosylmethionine analogue: a tool for the functional analysis of methyltransferases. J Am Chem Soc. 2010;132(11):3642–3.

Levy-Sakin M, Ebenstein Y. Beyond sequencing: optical mapping of DNA in the age of nanotechnology and nanoscopy. Curr Opin Biotechnol. 2013;24(4):690–8.

Liutkevičiūtė Z, Kriukienė E, Grigaitytė I, Masevičius V, Klimašauskas S. Methyltransferase-directed derivatization of 5-hydroxymethylcytosine in DNA. Angew Chem Int Ed Engl. 2011;50(9):2090–3.

Liutkevičiūtė Z, Kriukienė E, Ličytė J, Rudytė M, Urbanavičiūtė G, Klimašauskas S. Direct decarboxylation of 5-carboxylcytosine by DNA C5-methyltransferases. J Am Chem Soc. 2014;136(16):5884–7.

Liutkevičiūtė Z, Lukinavičius G, Masevičius V, Daujotytė D, Klimašauskas S. Cytosine-5-methyltransferases add aldehydes to DNA. Nat Chem Biol. 2009;5(6):400–2.

Lukinavičius G, Lapienė V, Stasevskij Z, Dalhoff C, Weinhold E, Klimašauskas S. Targeted labeling of DNA by methyltransferase-directed transfer of activated groups (mTAG). J Am Chem Soc. 2007;129(10):2758–9.

Lukinavičius G, Lapinaitė A, Urbanavičiūtė G, Gerasimaitė R, Klimašauskas S. Engineering the DNA cytosine-5 methyltransferase reaction for sequence-specific labeling of DNA. Nucleic Acids Res. 2012;40(22):11594–602.

Lukinavičius G, Tomkuvienė M, Masevičius V, Klimašauskas S. Enhanced chemical stability of adomet analogues for improved methyltransferase-directed labeling of DNA. ACS Chem Biol. 2013;8(6):1134–9.

Mai V, Comstock LR. Synthesis of an azide-bearing N-mustard analogue of S-adenosyl-L-methionine. J Org Chem. 2011;76(24):10319–24.

Masevičius V, Nainytė M, Klimašauskas S. Synthesis of S-adenosyl-L-methionine analogs with extended transferable groups for methyltransferase-directed labeling of DNA and RNA. Curr Protoc Nucleic Acid Chem. 2016;64:1.36.1–1.36.13.

Motorin Y, Burhenne J, Teimer R, Koynov K, Willnow S, Weinhold E, et al. Expanding the chemical scope of RNA:methyltransferases to site-specific alkynylation of RNA for click labeling. Nucleic Acids Res. 2011;39(5):1943–52.

Neely RK, Dedecker P, Hotta JI, Urbanavičiūtė G, Klimašauskas S, Hofkens J. DNA fluorocode: a single molecule, optical map of DNA with nanometre resolution. Chem Sci. 2010;1: 453–60.

Pastor WA, Pape UJ, Huang Y, Henderson HR, Lister R, Ko M, et al. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature. 2011;473(7347):394–7.

Peters W, Willnow S, Duisken M, Kleine H, Macherey T, Duncan KE, et al. Enzymatic site-specific functionalization of protein methyltransferase substrates with alkynes for click labeling. Angew Chem Int Ed Engl. 2010;49(30):5170–3.

Pignot M, Siethoff C, Linscheid M, Weinhold E. Coupling of a Nucleoside with DNA by a Methyltransferase. Angew Chem Int Ed Engl. 1998;37(20):2888–91.

Pljevaljčić G, Pignot M, Weinhold E. Design of a new fluorescent cofactor for DNA methyltransferases and sequence-specific labeling of DNA. J Am Chem Soc. 2003;125(12):3486–92.

Pljevaljčić G, Schmidt F, Scheidig AJ, Lurz R, Weinhold E. Quantitative labeling of long plasmid DNA with nanometer precision. Chembiochem. 2007;8(13):1516–9.

Pljevaljčić G, Schmidt F, Weinhold E. Sequence-specific methyltransferase-induced labeling of DNA (SMILing DNA). Chembiochem. 2004;5(3):265–9.

Plotnikova A, Osipenko A, Masevičius V, Vilkaitis G, Klimašauskas S. Selective covalent labeling of miRNA and siRNA duplexes using HEN1 methyltransferase. J Am Chem Soc. 2014;136(39):13550–3.

Prescher JA, Bertozzi CR. Chemistry in living systems. Nat Chem Biol. 2005;1(1):13–21.

Ramadan M, Bremner-Hay NK, Carlson SA, Comstock LR. Synthesis and evaluation of N6-substituted azide- and alkyne-bearing N-mustard analogs of S-adenosyl-l-methionine. Tetrahedron. 2014;70(34):5291–7.

Schlenk F, Dainko JL. The S-n-propyl analogue of S-adenosylmethionine. Biochim Biophys Acta. 1975;385(2):312–23.

Schmidt FH, Huben M, Gider B, Renault F, Teulade-Fichou MP, Weinhold E. Sequence-specific Methyltransferase-Induced Labelling (SMILing) of plasmid DNA for studying cell transfection. Bioorg Med Chem. 2008;16(1):40–8.

Schulz D, Holstein JM, Rentmeister A. A chemo-enzymatic approach for site-specific modification of the RNA cap. Angew Chem Int Ed Engl. 2013;52(30):7874–8.

Schumacher A, Kapranov P, Kaminsky Z, Flanagan J, Assadzadeh A, Yau P, et al. Microarray-based DNA methylation profiling: technology and applications. Nucleic Acids Res. 2006;34(2):528–42.

Serva S, Lagunavičius A. Direct conjugation of peptides and 5-hydroxymethylcytosine in DNA. Bioconjug Chem. 2015;26(6):1008–12.

Singh S, Zhang J, Huber TD, Sunkara M, Hurley K, Goff RD, et al. Facile chemoenzymatic strategies for the synthesis and utilization of S-adenosyl-(L)-methionine analogues. Angew Chem Int Ed Engl. 2014;53(15):3965–9.

Song CX, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, et al. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat Biotechnol. 2011;29(1):68–72.

Stecher H, Tengg M, Ueberbacher BJ, Remler P, Schwab H, Griengl H, et al. Biocatalytic Friedel-Crafts alkylation using non-natural cofactors. Angew Chem Int Ed Engl. 2009; 48(50):9546–8.

Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet. 2008;9(6):465–76.

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2009;324(5929):930–5.

Teague B, Waterman MS, Goldstein S, Potamousis K, Zhou S, Reslewic S, et al. High-resolution human genome structure by single-molecule analysis. Proc Natl Acad Sci U S A. 2010;107(24):10848–53.

Tomkuvienė M, Clouet-d'Orval B, Černiauskas I, Weinhold E, Klimašauskas S. Programmable sequence-specific click-labeling of RNA using archaeal box C/D RNP methyltransferases. Nucleic Acids Res. 2012;40(14):6765–73.

Townsend AP, Roth S, Williams HEL, Stylianou E, Thomas NR. New S-Adenosyl-l-methionine Analogues: synthesis and reactivity studies. Org Lett. 2009;11(14):2976–9.

Vranken C, Deen J, Dirix L, Stakenborg T, Dehaen W, Leen V, et al. Super-resolution optical DNA Mapping via DNA methyltransferase-directed click chemistry. Nucleic Acids Res. 2014;42(7):e50.

Wang R, Islam K, Liu Y, Zheng W, Tang H, Lailler N, et al. Profiling genome-wide chromatin methylation with engineered posttranslation apparatus within living cells. J Am Chem Soc. 2013;135(3):1048–56.
Wang R, Zheng W, Yu H, Deng H, Luo M. Labeling substrates of protein arginine methyltransferase with engineered enzymes and matched S-adenosyl-L-methionine analogues. J Am Chem Soc. 2011;133(20):7648–51.
Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet. 2005;37(8):853–62.
Weier HU, Wang M, Mullikin JC, Zhu Y, Cheng JF, Greulich KM, et al. Quantitative DNA fiber mapping. Hum Mol Genet. 1995;4(10):1903–10.
Weller RL, Rajski SR. DNA methyltransferase-moderated click chemistry. Org Lett. 2005;7(11):2141–4.
Weller RL, Rajski SR. Design, synthesis, and preliminary biological evaluation of a DNA methyltransferase-directed alkylating agent. Chembiochem. 2006;7(2):243–5.
Wilkinson S, Diechtierow M, Estabrook RA, Schmidt F, Huben M, Weinhold E, et al. Molecular scale architecture: engineered three- and four-way junctions. Bioconjug Chem. 2008;19(2):470–5.
Willnow S, Martin M, Luscher B, Weinhold E. A selenium-based click AdoMet analogue for versatile substrate labeling with wild-type protein methyltransferases. Chembiochem. 2012;13(8):1167–73.
Winter JM, Chiou G, Bothwell IR, Xu W, Garg NK, Luo M, et al. Expanding the structural diversity of polyketides by exploring the cofactor tolerance of an inline methyltransferase domain. Org Lett. 2013;15(14):3774–7.
Zhang C, Weller RL, Thorson JS, Rajski SR. Natural product diversification using a non-natural cofactor analogue of S-adenosyl-L-methionine. J Am Chem Soc. 2006;128(9):2760–1.
Zhang L, Szulwach KE, Hon GC, Song CX, Park B, Yu M, et al. Tet-mediated covalent labelling of 5-methylcytosine for its genome-wide detection and sequencing. Nat Commun. 2013;4:1517.
Zohar H, Muller SJ. Labeling DNA for single-molecule experiments: methods of labeling internal specific sequences on double-stranded DNA. Nanoscale. 2011;3(8):3027–39.