Ane Sesma · Tobias von der Haar *Editors*

Fungal RNA Biology

Springer

Fungal RNA Biology

Ane Sesma · Tobias von der Haar
Editors

Fungal RNA Biology

Springer

Editors
Ane Sesma
Centre for Plant Biotechnology
and Genomics
Universidad Politécnica de Madrid
Pozuelo de Alarcón, Madrid
Spain

Tobias von der Haar
Kent Fungal Group, School of Biosciences
University of Kent
Kent
UK

ISBN 978-3-319-05686-9 ISBN 978-3-319-05687-6 (eBook)
DOI 10.1007/978-3-319-05687-6
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014936619

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Eukaryotic gene expression is tightly controlled at multiple levels and coordinates the expression of proteins governing related biological processes, i.e. cellular growth and differentiation. Some of the molecular mechanisms controlling these gene networks are conserved among eukaryotes while others are absent or have evolved differently. This generates interesting diversity from the point of view of biologists trying to understand life, as well as challenging diversity from the point of view of those who attempt to control biology for the improvement of health, the environment or the economy.

It is estimated that more than one million fungal species exist on Earth. These include organisms of great importance for industry, medicine and agriculture. Despite their diversity in live styles (free-living, saprophytic, parasitic and mutualistic), fungi share common features distinctive from plants and animals and have been grouped taxonomically as an independent eukaryotic kingdom. Currently, more than 250 fungal species have been sequenced and this number will grow exponentially in the coming decade, thanks to efforts such as the 1,000 Fungal Genomes Project (http://1000.fungalgenomes.org/). Although the knowledge generated at a molecular level is clearly ahead on unicellular fission and budding yeasts, filamentous fungi represent invaluable tools for understanding additional features of complex eukaryotes such as multicellular development, pathogenesis, natural product synthesis, small RNA-mediated gene silencing, DNA methylation and programmed cell death. Their manageable genome sizes and vast diversity provide excellent tools and helpful insights to understand common and new regulatory mechanisms of gene expression, including their evolutionary perspective.

This advanced book on "Fungal RNA Biology" is a reflection of the work of many talented colleagues and individuals, whose collaborative efforts have made this project possible. We are really grateful to the authors for the effort involved. This book tries to cover the most relevant aspects and groundbreaking studies over the recent years on RNA-mediated mechanisms in model unicellular yeasts and filamentous fungi. Fifteen chapters, written by experts in their fields, describe the RNA-dependent processes that take place in a fungal cell, ranging from formation of coding and non-coding RNAs to mRNA splicing, export, localisation, degradation and translation. Other aspects related to RNA metabolism and gene expression are also covered such as ribosomal RNA biogenesis, gene silencing,

involvement of tRNA modifications in protein synthesis and heterochromatin regulation. Two methodological chapters based on biochemical and bioinformatic approaches conclude the book. This book highlights the commonalities and particularities of the fungal RNA machinery and RNA-dependent processes with higher eukaryotes, including remaining questions and future challenges in this area. Accordingly, it is a valuable resource for students and researchers studying RNA-dependent processes.

Contents

About the Editors

Dr. Ane Sesma obtained her B.Sc. in Chemistry, majoring in Biochemistry and Molecular Biology (Universidad Autónoma de Madrid, Spain, June 1994). During her Ph.D., she worked on origin of replications of native plasmids and further involvement in *Pseudomonas syringae* virulence. She expanded her expertise in plant–fungal interactions investigating mechanism of pathogenesis in the rice blast fungus *Magnaporthe oryzae* (The Sainsbury Laboratory, Norwich, UK, 2001). In July 2005, she started her independent research programme after being awarded a BBSRC David Phillips fellowship at the John Innes Centre (Norwich, UK). The award allowed her to investigate new aspects of fungal pathogenesis and RNA metabolism (pre-mRNA 3′ end processing). She joined the Centre for Plant Biotechnology and Genomics (CBPG-Technical University of Madrid, Spain) as an assistant professor in September 2011.

Further information about the Sesma lab: http://www.cbgp.upm.es/en/riceblastfungus.php.

Dr. Tobias von der Haar obtained a Diploma degree in Biology (University of Bielefeld, Germany, 1995). During his Ph.D., he worked on the role of the cap-binding protein eIF4E in regulating translation in baker's yeast. In 2004, he established his research group under a Research Career Development Fellowship from the Wellcome Trust (UK), with research interests in the areas of translation termination and the maintenance of translational fidelity. He presently works as Senior Lecturer at the University of Kent in Canterbury, UK, using combinations of computer modelling and experimental work to address how the network of ribosomes, mRNAs and tRNAs produces a functional and robust translational machinery in eukaryotic cells.

Further information about the von der Haar lab: http://yeaki.org/.

Chapter 1
RNA Polymerase II-Dependent Transcription in Fungi and Its Interplay with mRNA Decay

Xenia Peñate and Sebastián Chávez

Abstract Messenger RNA transcription in fungi, particularly in the budding yeast *Saccharomyces cerevisiae*, is one of the main models for transcriptional research. In this chapter, we review the main mechanisms that operate during fungal RNA polymerase II-dependent transcription, from the initiation step to the termination one. In the elongation phase, processing of the nascent transcript, including 5′ capping, splicing, 3′ end formation and transport, is coupled to transcription. The RNA polymerase II template is not naked DNA, but chromatin. We review the impact of chromatin in the elongation phase and in the phenomenon of RNA polymerase II backtracking. Strikingly, synthesis and degradation have been shown recently to be connected, resulting in a general buffering system for messenger RNA concentration. In this way, messenger RNA synthesis, processing, and degradation are interlinked and have the potential to influence each other.

Contents

X. Peñate · S. Chávez (✉)
Instituto de Biomedicina de Sevilla, Universidad de Sevilla-CSIC-Hospital Virgen del Rocío, Departamento de Genética, Universidad de Sevilla, Seville, Spain
e-mail: schavez@us.es

A. Sesma and T. von der Haar (eds.), *Fungal RNA Biology*,
DOI: 10.1007/978-3-319-05687-6_1,

Transcription Initiation

In fungi, as in all eukaryotes, three different nuclear RNA polymerases transcribe the genome. RNA polymerase I (RNA pol I) transcribes ribosomal DNA, RNA polymerase II (RNA pol II) is in charge of messenger RNA (mRNA) and some small nuclear RNA (snRNA), while RNA polymerase III synthesises transfer RNA (tRNA) and some other short RNAs (Vannini and Cramer 2012). RNA pol II is by far the best studied of them all and is the subject of this chapter. Nevertheless, all three nuclear transcriptional machineries share a conserved core and a basic mechanism of action (Vannini and Cramer 2012). The structure of RNA pol I, recently solved at high resolution, confirms this similarity, although it also uncovers some striking differences with RNA pol II (Engel et al. 2013; Fernandez-Tornero et al. 2013). Unless otherwise stated, details of the mechanisms of transcription refer to those in *Saccharomyces cerevisiae* because this is the best-known system.

The transcription unit is typically composed of the following: promoter, transcription start site (TSS), transcribed region, and transcription termination site (TTS). In the promoter, two regions of special importance are normally present: the TATA or TATA-like box and the upstream activating sequence (UAS) (Rhee and Pugh 2012; Guarente 1988). For an mRNA, the transcribed region can be divided into a 5′ untranslated region (5′ UTR), an open reading frame (ORF), and a 3′ untranslated region (3′ UTR). ORFs can be interrupted by introns, which are usually shorter in fungi than in metazoa (Kupfer et al. 2004). Intron density is highly variable in fungi, and ranges from 0.05 introns per gene in *S. cerevisiae* to 5.5 introns per gene in *Cryptococcus neoformans* (Engel et al. 2013).

Of the three phases in which we divide transcription (initiation, elongation, and termination), initiation has been long since recognized to be a rate-limiting and regulated process (Tjian 1978; Klages and Strubin 1995). Initiation can be further divided into preinitiation complex (PIC) assembly, promoter DNA melting and early initiation events (Cramer 2004).

PIC Assembly

The current view of PIC assembly involves sequence recognition by specific transcription factors (STFs), recruitment of accessory factors such as chromatin remodellers and coactivators and finally recruitment of RNA pol II and general transcription factors (GTFs) (Fig. 1.1a, b) (Venters and Pugh 2009b).

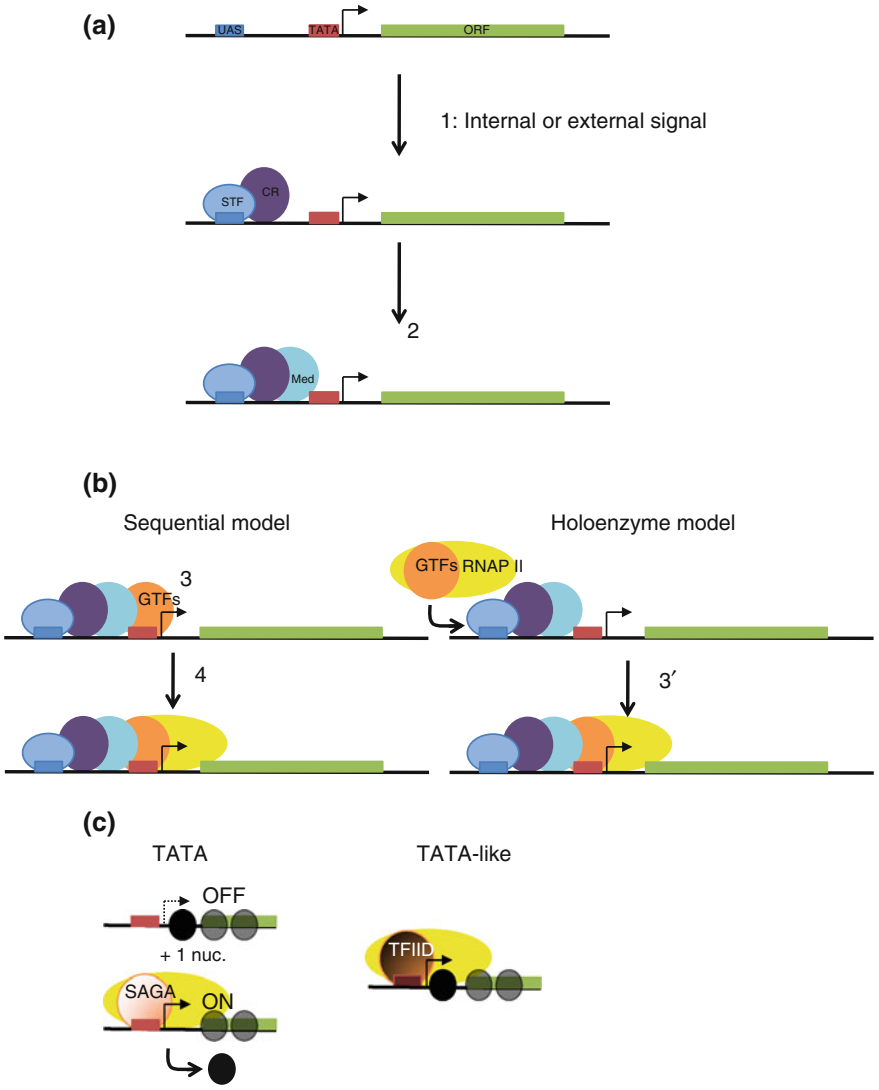


Fig. 1.1 Transcription initiation by RNA pol II. **a** Recruitment of the transcription machinery to a particular promoter starts with the recognition of specific sequences by STFs in response to a signal. STFs then recruit chromatin remodelling complexes (CR) and Mediator (Med); **b** RNA pol II and GTFs might be recruited in sequence (sequential model) or together (holoenzyme model); **c** Minor differences in the TATA element might determine a different initiation step in terms of co-activator recruitment (SAGA or TFIID) and chromatin dynamics (see the text for details)

STFs bind the UAS in response to stimuli. Thus, they are the link between signal transduction pathways and cellular output in transcription terms (Venters and Pugh 2009b). This sequence-specific binding is required for the induction of

gene expression, although some basal transcription can occur in its absence (Klages and Strubin 1995; Pugh 2000). Typically, the UAS is several hundred base pairs (bp) upstream from the TSS (Harbison et al. 2004), so it does not determine the exact TSS. GTFs perform this function (see below). STFs regulate transcription by recruiting chromatin remodelling complexes, GTFs, and co-activators (e.g. Mediator; Venters and Pugh 2009b).

Chromatin Remodelling Complexes

The organisation of eukaryotic DNA into nucleosomes is a physical constraint to transcription (Kornberg 2007). As a consequence, RNA pol II needs chromatin remodelling complexes to deal with chromatin. Promoters tend to be positioned in a nucleosome-free region (NFR), bracketed by the −1 and the +1 nucleosomes (Yuan et al. 2005), which makes these sequences more accessible to the binding of STFs and GTFs. NFRs are formed by the combination of anti-nucleosomal DNA sequences, chromatin remodelling activities, and, in some cases, transcription factor binding (Struhl and Segal 2013).

Chromatin remodelling complexes are ATPase motors that promote the destabilisation of DNA-histone interactions. Since some of these complexes are known to act independently of transcription (Lorch et al. 2006), their recruitment to the promoter region by STFs may be important for the formation of the NFR. Chromatin remodelling complexes may not only contribute to NFRs but also to −1 nucleosome removal, another way of contributing to RNA pol II recruitment to the PIC (Venters and Pugh 2009a). In other cases, chromatin remodelling complexes may repress transcription by creating a closed chromatin configuration (Fazzio et al. 2001).

GTFs

GTFs are a series of protein complexes, known as TFIIA, B, D, E, F and H that help the polymerase through the different initiation steps (Thomas and Chiang 2006). The TATA box is the starting DNA contact of the GTFs, and is located at a fixed, short distance from the TSS (60 bp) (Kornberg 2007; Venters and Pugh 2009b). It is recognised by TBP (TATA-binding protein), a subunit of TFIID (Papai et al. 2011). TFIID is also composed of at least 13 TBP-associated factors (TAFs) (Papai et al. 2011). Most TAFs are also present in an alternative complex called SAGA (Spt-Ada-Gcn5-acetyltransferase; Rodriguez-Navarro, 2009). SAGA is recruited to promoters with a canonical TATA box, normally at inducible genes. TFIID-bound promoters tend to be those of housekeeping genes, and their TATA box is somehow degenerated, so they are known as TATA-like elements (Rhee and Pugh, 2012). Along with TBP, the histone-modifying activities present in SAGA and TFIID are thought to help PIC formation (Grant et al., 1997; Thomas and Chiang, 2006). Among them, acetylation of the lysine residues present in histone tails is essential for promoter activation (Bhaumik, 2011) since acetylated histones are recognised

by the bromodomains present in nucleosome remodelling complexes like SWI-SNF and RSC (Chatterjee et al. 2011). Reciprocally, deacetylation of histone tails plays a role in promoter repression (Takahata et al. 2009; Mitra et al. 2006).

DNA is bent around the polymerase, thanks to TBP binding to the TATA box (Kornberg 2007). Two other GTFs, TFIIA and TFIIB, stabilise TBP-DNA interactions (Thomas and Chiang 2006). What makes the polymerase start at the TSS specifically? First of all, the concept of a single TSS for each gene is currently being re-examined due to the finding of many alternative transcript isoforms genome-wide (Pelechano et al. 2013). Nonetheless, it is clear that the polymerase does not start just anywhere. Structural data suggest that the interaction between TBP, TFIIB and the polymerase determine the start site, with help from TFIIF (Vannini and Cramer 2012; Kornberg 2007; Cramer 2004). Recent data suggest that start site selection is dependent on the gene type, classifying genes as TATA or TATA-like containing. According to these data, the position of both the TATA-like box and the +1 nucleosome define the TSS of the TATA-like containing genes. The TSS of the TATA-containing genes is instead located further downstream, which leaves room for the polymerase to scan the DNA for a TSS. These latter genes may require the eviction of the +1 nucleosome to start transcription (Fig. 1.1c; Rhee and Pugh 2012; Struhl and Segal 2013), which may explain why their nucleosome architecture is not canonical (Albert et al. 2007). TFIIF helps TSS selection by preventing non-specific DNA binding, and it is required for PIC stability (Cramer 2004; Vannini and Cramer 2012).

Mediator was not initially considered a GTF, but is required for the transcription of most RNA pol II promoters (Kornberg 2007). It is a large protein complex that acts as a bridge between STFs and the PIC (Kornberg 2007).

RNA Pol II

There are two models to explain how RNA pol II binds promoter DNA (Fig. 1.1b; Cramer 2004). In both of them, however, the result is the binding of not only RNA pol II, but also a number of GTFs required for initiation (Cramer 2004; Kornberg 2007). In the sequential model, GTFs bind promoter DNA in a specific order, and they recruit RNA pol II (Cramer 2004; Thomas and Chiang 2006). This view is mainly supported by in vitro transcription studies (Thomas and Chiang 2006). In the holoenzyme model, RNA pol II and GTFs are bound in solution and they bind promoter DNA together (Cramer 2004; Thomas and Chiang 2006). Recent evidence points to a chimera of these two theories: every GTF except TFIIH seems to be present at certain promoters, to which the holoenzyme formed by RNA pol II and TFIIH can be recruited (Venters and Pugh 2009b).

RNA pol II is composed of 12 subunits, all of which are required for initiation (Cramer 2004). Except Rpb4 and Rpb9, they are all essential in *S. cerevisiae*. Interestingly, Rpb4 is essential in *Schyzosaccharomyces pombe* (Cramer 2004), which emphasises that there might be substantial differences in transcription within the fungal kingdom. The two largest subunits, Rpb1 and Rpb2, form a positively charged "cleft"

that binds DNA and contains the active centre (Cramer 2004). A region in Rpb1, called the trigger loop, acts as a selective door that traps the correctly matched NTP at the active site (Kornberg 2007). Yet another Rpb1 region, the bridge helix, is required for both translocation and DNA-RNA duplex binding (Kornberg 2007).

In all eukaryotes, Rpb1 contains an essential C-terminal domain that is composed of multiple repeats of the consensus sequence YSPTSPS (Cramer 2004). This domain, known as the CTD, is thought to be extended when phosphorylated (possibly covering long distances from the active centre) and compacted when unphosphorylated (Cramer 2004). All the residues within the repeat, except prolines, can be phosphorylated, and prolines can be isomerised. The combination of post-translational modifications in the CTD is known as the CTD code (Jeronimo et al. 2013). The additive effects of specific kinases, phosphatases and prolyl isomerases (Jeronimo et al. 2013) generate this code. The CTD must be unphosphorylated for the polymerase to be recruited to promoter DNA (Cramer 2004). Then, the CTD is phosphorylated in Ser5 during promoter escape and later in Ser2 during elongation. These phosphorylation events are likely to be dependent on the regulatory features of genes since considerable gene-specificity has been observed for CTD phosphorylation across the *S. cerevisiae* genome (Kim et al. 2010). The functional consequences of the CTD code will be further discussed later in this chapter. The length of the CTD increases through evolution, and is maximal in animals with more than 50 repeats (Liu et al. 2010). In fungi, the number of repeats ranges between 15 in some microsporidians to more than 30 in some basidiomycetes (Liu et al. 2010). Interestingly, the conservation of the heptapeptide sequence is heterogeneous among fungi, and ascomycete yeasts are more similar to animals and plants than to filamentous ascomycetes like *Aspergillus* or *Penicillium* (Liu et al. 2010).

Promoter DNA Melting

RNA pol II cannot melt DNA itself, although it can maintain an open transcription bubble (Cramer 2004). To melt DNA, TFIIE and TFIIH are required (Cramer 2004). TFIIE recruits TFIIH and regulates its activity (Venters and Pugh 2009b), and it may bind the non-template strand after both strands have been separated (Grunberg and Hahn 2013). The exact DNA melting mechanism remains unknown. Based on current data, two models are possible, depending on the conformation of the clamp. The clamp is a region in RNA pol II formed by Rpb1 and Rpb2. RNA pol II crystals can exist in two different states: with an open clamp or with a closed clamp. The open clamp allows duplex DNA binding, but the closed clamp can only accommodate a single strand (Cramer 2004; Grunberg and Hahn 2013). Therefore, the question is whether the DNA melts on the surface or inside the clamp. In one model, duplex DNA may bind the polymerase surface, be melted there, and the template strand would relocalise to enter the closed clamp (Cramer 2004). In the alternative model, duplex DNA would localise into the open clamp and be melted there. However, this last model seems more unlikely since

Rpb4/7 forms a barrier in the polymerase structure that does not allow the open clamp state or, consequently, the entrance of the duplex DNA (Cramer 2004). Interestingly, recent data suggest that TFIIF may help in the opening of the clamp (Grunberg and Hahn 2013).

Together with the exact location, the enzymatic activity responsible for DNA melting is also unknown. Recently, it has been suggested that DNA melting can be the result of DNA translocation into the polymerase (Grunberg and Hahn 2013). Downstream of the PIC, a subunit of TFIIH would insert DNA into the polymerase, thus creating a torsional stress, since the position of the PIC is fixed, thanks to TBP binding. Duplex DNA would melt to relieve the torsional stress. This is a new and an attractive hypothesis about DNA melting that need to be tested. It will be interesting to see whether it can help discriminate between the two models discussed above.

Promoter Escape

The structure formed by the PIC and the melted template DNA is called the open complex (Grunberg and Hahn 2013). Once the open complex is formed, DNA-dependent RNA polymerisation starts. The TFIIB structure suggests that it stabilises the RNA-DNA interaction of the nascent transcript up to a few nucleotides (Grunberg and Hahn 2013).

Short RNAs of up to 15 nucleotides are released from the early transcribing complex (Cramer 2004). Abortive initiation is an intrinsic characteristic of RNA synthesis by RNA pol II (Cramer 2004). Overcoming this crucial step is known as promoter clearance (Cramer 2004). TFIIE and TFIIH help the polymerase at this step (Cramer 2004). In particular, phosphorylation of the CTD at Serine 5 (Ser5) by Kin28, a subunit of TFIIH, helps the dissociation of the polymerase from the rest of the PIC, thus enhancing promoter escape (Sogaard and Svejstrup 2007). Promoter escape also requires the dissociation of TFIIB from the polymerase (Vannini and Cramer 2012), although it is still not known whether this is related to P-Ser5.

Transcription Elongation

The best-known gene expression mechanisms regulate transcription during pre-initiation complex assembly and initiation, but a significant number of regulatory programs operate in post-initiation steps (Margaritis and Holstege 2008). Accordingly, elongation is not only the step that immediately follows promoter escape, but it is also an additional opportunity to regulate transcription. Regulation at the transcription elongation level is a widespread phenomenon in metazoa (Brannan and Bentley 2012), and it is also an important ingredient of the regulatory cocktail in fungi (Gomez-Herreros et al. 2012a).

The CTD Code Through Elongation

The CTD of RNA pol II acts as the control panel that leads the behaviour of the enzyme during elongation and reflects its molecular state.

The phosphorylation state of the CTD changes as the polymerase advances through the gene (Fig. 1.2a) within a CTD cycle with numerous functional implications. Each serine in the heptapeptide repeat is phosphorylated with a particular pattern that is dependent on the absolute distance from the TSS and to the TTS (Bataille et al. 2012). Generally, the CTD can be viewed as a tuneable platform to which distinct factors are recruited at different times within the transcription cycle (Jeronimo et al. 2013). This timely recruitment is important for transcription-coupled processes (see "Transcription-Coupled Processes and Termination"), and for transcription itself.

Soon after initiation, Kin28 phosphorylates Ser5. This modification allows: (1) promoter escape (see above); (2) the binding of capping and splicing factors (see "Transcription-Coupled Processes and Termination"); and (3) the binding of histone methyltransferases and histone deacetylases (Jeronimo et al. 2013). A complex sequence of events (including the phosphorylation and ubiquitylation of several factors, and methylation of histones) starts with Ser5 phosphorylation and ends with histone acetylation around the promoter (Venters and Pugh 2009b). Acetylated histones are negatively charged and less strongly bound to DNA and are, hence, a better landscape for a transcribing polymerase. Histone deacetylases counteract the effect of promoter histone acetylation in the transcribed region, thus avoiding cryptic initiation (Jeronimo et al. 2013).

Ctk1 is the major Ser2 kinase. Ser2 phosphorylated CTD recruits other histone methyltransferases and activates histone deacetylases (Jeronimo et al. 2013). Therefore, the CTD is the most important link among transcription, RNA processing and chromatin modification (see later).

Pausing and Backtracking

During undisturbed elongation, RNA pol II can bind the DNA–RNA hybrid and translocate. Elongation complex stability depends on the binding of the polymerase to the hybrid. Therefore, this binding needs to be strong, but not so strong as to avoid translocation (Cramer 2004). Biochemical and structural data indicate that translocation may involve conformational changes within the polymerase (Cramer 2004). According to the classic paradigm, transcription elongation is a uniform process in which RNA pol II advances timely from initiation to elongation with no relevant perturbations. The distribution of RNA pol II along a gene in a population of cells is gene-specific in *S. cerevisiae* (Rodriguez-Gil et al. 2010; Churchman and Weissman 2011), recapitulating similar observations in higher eukaryotes (Core et al. 2008; Gilchrist et al. 2009) (see below). A similar

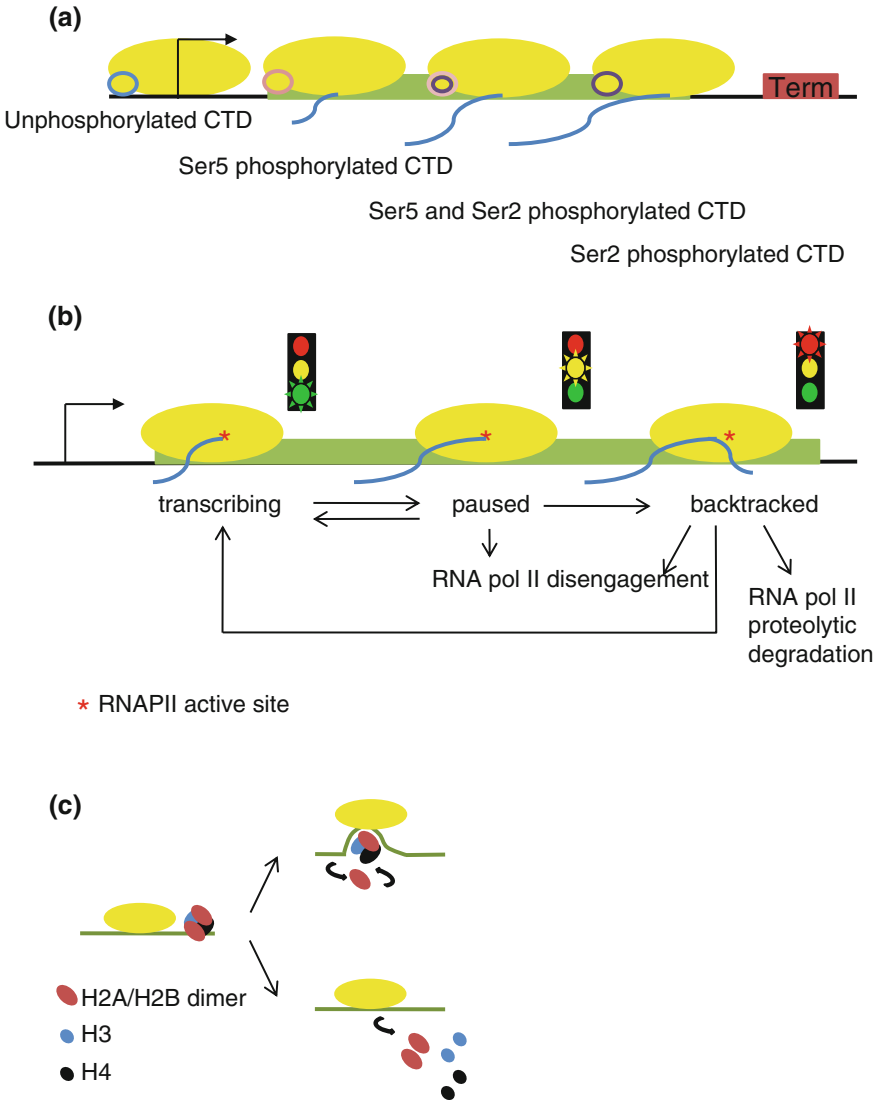


Fig. 1.2 Transcription elongation by RNA pol II. **a** CTD phosphorylation changes along the elongation process. **b** Different obstacles, mainly nucleosomes, can make the polymerase pause or backtrack during elongation. *Arrows* summarise the main processes that solve every situation. **c** Proposed alternative modes of transcription through chromatin

conclusion has been reached in single-cell experiments, in which RNA pol II progression along a transcribed gene has proved to be a discontinuous process that combines short advances with pauses of a diverse time length (Darzacq et al. 2007). Therefore, different methodologies convey pausing as a frequent phenomenon during RNA pol II-dependent transcription.

Evidence from in vitro experiments has shown that RNA pol II pausing is highly unstable, and results in either forward transcription or stable arrest (Gu and Reines 1995). Moreover, it is now common understanding that RNA pol II elongates not in a unidirectional way, but by oscillating between forward and backward movements. DNA–RNA hybrid stability normally enhances forward movement, but any hindrance of the polymerase forward movement can increase the chance of backtracking (see below) (Sigurdsson et al. 2010; Fig. 1.2b).

RNA pol II backtracking involves a reverse movement that produces loss of contact between the RNA 3′ end and the active site of the enzyme (Cheung and Cramer 2011). As a consequence of this relocalisation of the 3′ end, backtracked RNA polymerases cannot incorporate new nucleotides.

The backtracked RNA pol II complex is extremely stable due to the strong binding of the backtracked RNA to the funnel domain (Cheung and Cramer 2011). This domain is highly conserved from fungi to higher eukaryotes, suggesting that RNA pol II backtracking is anything but accidental during RNA pol II-dependent elongation (Cheung and Cramer 2011).

From the first phosphodiester bonds to the production of full-length mRNA, RNA pol II is bound to encounter a wide range of hindrances. In vivo and in vitro studies have described a number of such hindrances which make the polymerase pause and backtrack. One of the causes that brings about RNA pol II arrest is scarcity of nucleotides. Drugs provoking the depletion of NTP pools, like mycophenolic acid and 6-azauracil, decrease the processivity and elongation rate of RNA pol II in vivo (Mason and Struhl 2005), and increase the frequency of arrest in vitro (Powell and Reines 1996).

Another cause of RNA pol II arrest is the presence of physical obstacles in DNA. Nucleosomes, which cover almost the entire length of eukaryotic genomes, have been shown to promote backtracking in vitro. Accordingly, transcription of nucleosomal templates is stimulated by those factors that reactivate backtracked RNA pol II (Fish and Kane 2002) (see below). Topological constraints are another kind of hindrance potentially encountered by RNA pol II during elongation which can lead to stalling. RNA pol II-dependent elongation generates positive supercoiling, which needs to be removed by topoisomerases (Brill and Sternglanz 1988). Positive supercoiling is particularly relevant in yeast genes longer than 3 kb (Joshi et al. 2012). DNA lesions are the fourth kind of hindrance that can cause RNA polymerase arrest within transcribed regions. In most cases, this arrest is irreversible and needs to be solved by the degradation of the stalled polymerase (Daulny and Tansey 2009) (see below). Finally, pyrimidine-rich tracts can also favour RNA pol II arrest in vitro, even when the template takes a non-nucleosomal configuration (Izban and Luse 1991). Interestingly, there is a strong preference for pyrimidines in the interaction between backtracked RNA and the funnel (Cheung and Cramer 2011). This sequence preference for backtracking predicts the existence of an arresting code across the genome, which might explain the biased distribution of the polypyrimidine/polypurine tracts between template and non-template strands across the genome of *S. cerevisiae* (Brahmachari et al. 1997).

Factors Counteracting and Regulating RNA Pol II Arrest

Both paused and backtracked polymerases are enzymes engaged in DNA and not moving forward (Cheung and Cramer 2011). However, overall RNA polymerase pausing and arrest cannot be permanent in replicating nuclei since it would result in interference with replication forks, thus leading to genome instability, and eventually to cell death (Daulny and Tansey 2009).

While pausing can be solved without any external factor, backtracking often requires specific factors. Three different mechanisms have been described to solve RNA pol II backtracking: (1) RNA cleavage, which allows the RNA 3′ end to relocate at the active site; (2) reversion of the backtracked state by the polymerase forward movement; and (3) eviction of arrested RNA pol II by ubiquitylation-mediated degradation. The factors favouring these mechanisms are not essential for *S. cerevisiae* growth under standard culture conditions, but they are necessary when cultivated in the presence of NTP-depleting drugs (see above) (Gaillard et al. 2009).

RNA cleavage is catalysed by the endogenous endonucleolytic activity of RNA pol II, which becomes highly stimulated by cleavage factor TFIIS (Izban and Luse 1992). The structural interaction between the backtracked configuration of RNA pol II and TFIIS explains this stimulation (Cheung and Cramer 2011). The domain III of TFIIS enters RNA pol II through its pore domain and reaches the active site, thus enabling the displacement of RNA from the funnel. In addition, TFIIS complements the active site when catalysing the cleavage reaction (Cheung and Cramer 2011).

Enhancing the polymerase forward movement is an alternative way of solving backtracking. The Ccr4-Not complex interacts with the emerging transcript to favour the RNA pol II forward movement and, therefore, to promote the resumption of elongation without RNA cleavage (Kruk et al. 2011). In addition to these reactivation mechanisms, *Saccharomyces* cells can contend with irreversibly blocked transcriptional complexes by proteolytic degradation in an ubiquitylation-dependent manner (Somesh et al. 2005).

The comparison of the genomic patterns of run-on signals, reflecting elongation-proficient RNA polymerases, with the genomic distribution of total RNA pol II measured by ChIP is an indirect way of detecting backtracked RNA polymerases (Perez-Ortin et al. 2011). In general, a good correlation exists between the run-on and ChIP signals across the *Saccharomyces* genome, but some gene clusters exhibit lower run-on values than expected according to their ChIP signals (Pelechano et al. 2009). One particularly striking finding was the difference observed in the genes encoding ribosomal proteins (RP) (Pelechano et al. 2009). These and other results indicate RNA pol II backtracking to be an important element in the transcriptional regulation of the yeast genome [reviewed in (Gomez-Herreros et al. 2012a)].

In metazoa, the regulatory processes based on RNA pol II pausing and backtracking require the action of reactivation mechanisms, like that stimulated by TFIIS (Adelman et al. 2005). This is also the case in *S. cerevisiae*, where RP genes need TFIIS to fulfil the regulatory response of ribosome biogenesis under the transcriptional stress imposed by NTP-depleting drugs (Gomez-Herreros et al. 2012b).

The requirement of TFIIS under these conditions can be overcome by deleting some RP regulatory factors, like Sfp1 (Gomez-Herreros et al. 2012b). Surprisingly, the absence of Sfp1 provokes a general decrease in RNA pol II backtracking in many functionally unrelated genes, which indicates that backtracking is a highly regulated process that operates genome-wide (de Miguel and Chávez, unpublished results).

Chromatin Dynamics During Transcription Elongation

In vitro and in vivo experiments have shown that non-remodelled nucleosomes are powerful inhibitors of transcription during both initiation and elongation (see above). The chromatin landscape of genomes show that nucleosome density and positions vary vastly along fungal chromosomes (Tsankov et al. 2011). The nucleosome repeat of RP genes is shorter than that of standard genes (Weiner et al. 2010). Chromatin covalent modifications and histone variants also exhibit gene-to-gene variation across the genome (Zhang and Pugh 2011), while RP genes exhibit very peculiar histone marks (Peñate, Pozo and Chávez, unpublished). Differential chromatin dynamics is, therefore, a possible explanation for the diverse tendency of RNA pol II to backtrack across the genome. According to this hypothesis, some chromatin configurations would promote a higher frequency of RNA pol II backtracking than others.

In those genes exhibiting higher levels of RNA pol II backtracking, the factors that reactivate arrested RNA pol II like TFIIS would be in great demand. In vitro experiments have demonstrated that TFIIS is indeed required for efficient RNA pol II-dependent elongation through a nucleosome (Kireeva et al. 2005), and that it can synergise with other elongation factors to stimulate the RNA pol II traversal of a nucleosome (Luse et al. 2011).

The classical view of histone dynamics during transcription elongation assumes full histone eviction to be an absolute requirement before RNA pol II transcription. Supporting this view, full histone eviction is clearly demonstrated to occur in some very well-studied inducible genes of *Saccharomyces* (Schwabish and Struhl 2004, 2006). It involves nucleosome reassembly after RNA pol II passage. This cotranscriptional chromatin reassembly seems essential for controlling cryptic transcription within the gene bodies in the *Saccharomyces* genome (Cheung et al. 2008).

However, the in vitro nucleosome traversal by RNA pol II in the systems stimulated by TFIIS takes place without nucleosome displacement thanks to the formation of an intranucleosomal DNA loop (Ø loop) that contains the transcribing enzyme (Kulaeva et al. 2009). The results obtained by atomic force microscopy confirm that a nucleosome can survive transcription by RNA pol II without complete disassembly, but just the removal of a single H2A–H2B dimer (Bintu et al. 2011). The H2A–H2B chaperone Nap1 seems to facilitate the removal of the histone dimer during this conservative mode of nucleosome transcription without full histone eviction (Kuryan et al. 2012).

Therefore, does transcription need nucleosome eviction or not? The easiest way of reconciling all the results would be the existence of two alternative modes of transcribing chromatin. Under one mode, transcription elongation would involve the previous eviction of the whole histone octamer. This eviction-dependent mode of transcribing chromatin would not involve TFIIS, and we would expect a low level of backtracking. Under the alternative mode, a single histone H2A-H2B dimer would be evicted before elongation, to form the so-called hexasome (Arimura et al. 2012). It has been recently demonstrated that nucleosome survival during elongation requires specific RNA pol II-histone contacts (Chang et al. 2013). In the model proposed herein these frequent histone–RNA pol II contacts would involve more backtracking (Fig. 1.2c).

The existence of two alternative modes of handling chromatin during transcription elongation helps explain the gene specificity of RNA pol II backtracking, as we observed in *Saccharomyces*. Those genes exhibiting high backtracking frequency and highly dependent on TFIIS, like RP genes, would be dominated by the hexasome mode, whereas those genes with lower backtracking levels when transcribed would be dominated by the histone-eviction mode. The latter would be the case of SAGA-dependent inducible genes, which are poorly dependent on TFIIS and are less prone to backtracking (Pelechano et al. 2009; Gomez-Herreros et al. 2012b). The observation of positioned nucleosomal profiles being more resistant to transcription in RP genes, and in general in TATA-like genes, than in canonical TATA, SAGA-dependent genes, supports this model (Zhang and Pugh 2011).

Saccharomyces genetics is contributing to confirm this model. We have recently isolated a novel yeast factor that affects chromatin dynamics during elongation. Prefoldin, a complex so far known to be involved in cytoskeleton assembly, is required for full histone eviction during elongation (Millan-Zambrano et al. 2013). Histone levels after transcription induction are higher in prefoldin mutants than in the wild type. Otherwise, prefoldin mutants are healthy. The TFIIS mutant does not exhibit this defect in histone eviction. Interestingly, the double mutants lacking both prefoldin and TFIIS, are sick and exhibit generalised transcriptional defects (Millan-Zambrano et al. 2013). Hexasome and histone-eviction modes are, therefore, redundant pathways for transcription elongation through chromatin.

Transcription-Coupled Processes and Termination

Nascent RNA is subjected to a series of reactions that renders it a mature mRNA: 5′ capping, splicing, and 3′ end processing (Hsin and Manley 2012). Subsequently, mature mRNA is transported to the cytoplasm to be translated. All these processes are, at least for some genes, coupled to transcription, and in the case of 3′ end processing, the coupling is an absolute requirement for transcription termination.

5′ Capping

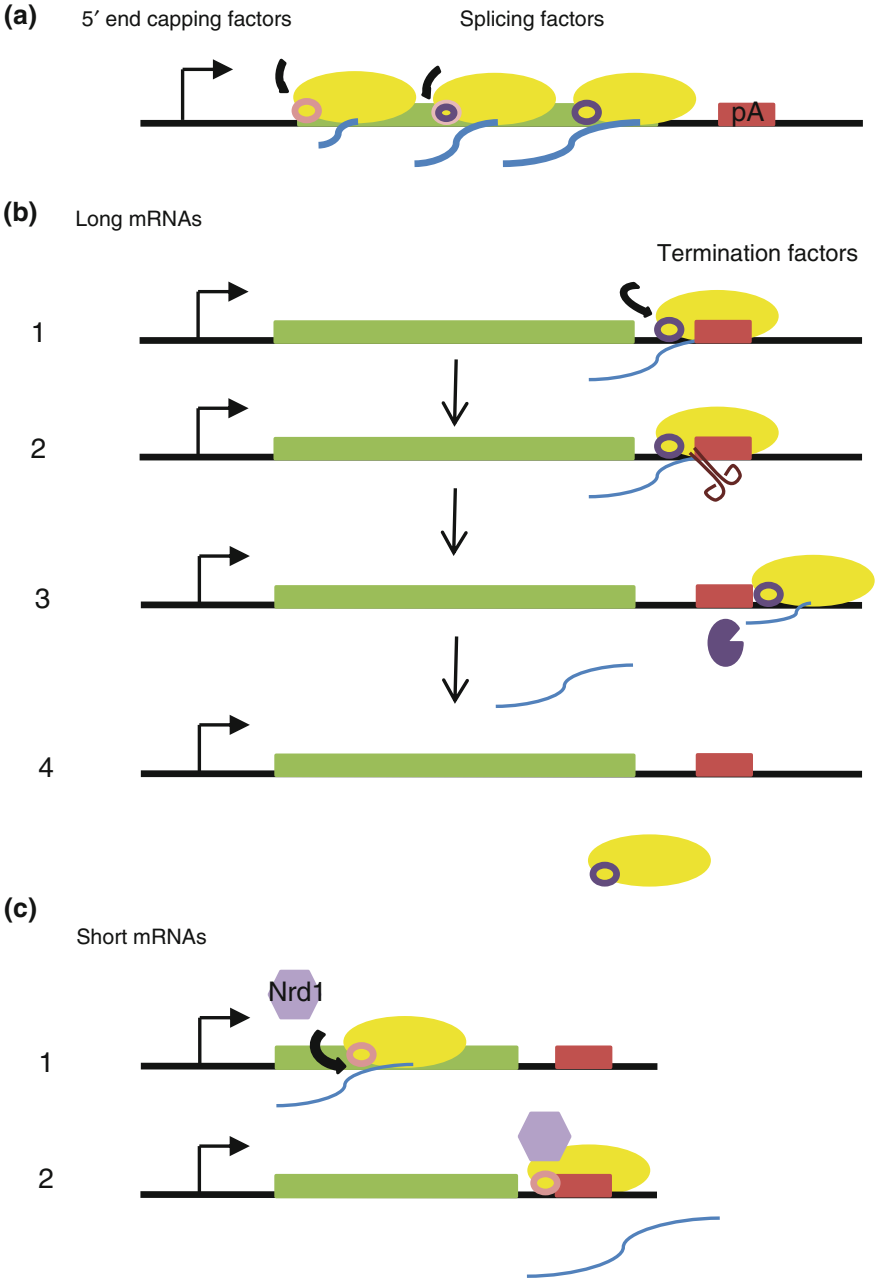
Mature mRNAs are modified on their 5′ end by a structure called cap. Cap formation involves three different activities: an RNA triphosphatase, a guanylyltransferase, and a methylase. Nascent RNA is only 25–30 nucleotides long when it is capped, so capping and transcription are co-occurring processes (Cramer 2004). In fact, the capping enzyme (guanylyltransferase) is recruited to P-Ser5 CTD (Fig. 1.3a) (Rodriguez et al. 2000; Venters and Pugh 2009b). This specificity explains why only RNA pol II transcripts are capped, since RNA pol I and III lack a CTD (Vannini and Cramer 2012). Tethering the capping enzyme to the CTD rescues the lethality of the substitution of all Ser5 in *S. pombe* (Hsin and Manley 2012), which indicates that quick 5′ capping is essential, and that it needs the physical link of the polymerase and the capping enzyme.

Splicing

Fungal genes have introns, regions that are transcribed but do not form part of mature mRNA (see above and Chap. 2 for more details). The process of removing these regions is called splicing. Although introns are scarce in *S. cerevisiae*, splicing has been extensively studied in this organism thanks to its simplicity. Here, as in other eukaryotes, splicing and transcription are linked physically and functionally. The CTD is, once again, the platform to which several splicing factors bind the polymerase (Hsin and Manley 2012). Indeed, *ctk1*Δ cells have low levels of Ser2-P CTD and show defects in splicing (Fig. 1.3a) (Phatnani et al. 2004). It has been proposed that the interaction of the splicing machinery with the CTD brings together the 5′ and 3′ ends of the intron, thus enhancing splicing efficiency (Morris and Greenleaf 2000). As transcription and splicing are linked, the elongation rate has an impact on splicing, and the opposite also seems to be true (Hsin and Manley 2012). Polymerase pausing can influence not only splicing efficiency, but also alternative splicing (Phatnani et al. 2004; Hsin and Manley 2012), although this last phenomenon is rare in fungi (Ast 2004).

3′ End Processing and Termination

The 3′ end of every mRNA is a polyA tail. This polyA tail is not encoded in DNA, but a polyadenylating enzyme, that forms part of the 3′ end processing machinery, adds the tail (Richard and Manley 2009; see Chap. 3 for more details). A single protein complex is necessary for both the transcription termination and 3′ end processing of nascent RNA: the cleavage and polyadenylation machinery (Venters and Pugh 2009b). In this way, these two processes are intrinsically linked.


(a)
5′ end capping factors
Splicing factors
pA
(b)
Long mRNAs
Termination factors
1
2
3
4
(c)
Short mRNAs
Nrd1
1
2

◀ **Fig. 1.3** Transcription-coupled processes and termination. **a** Both 5′ capping and splicing are coupled to transcription by the recruitment of the respective factors through an interaction with phosphorylated CTD. pA: poly(A) site. **b** Termination involves the recruitment of the cleavage and polyadenylation machinery upon the recognition of the TTS (*1*), RNA cleavage (*2*), 5′ to 3′ degradation of post-TTS RNA (*3*), and RNA pol II disassembly (*4*). The recruitment of the cleavage and polyadenylation machinery to most genes depends on the Ser2-P of the CTD (colour code as in Fig. 1.2). **c** The Nrd1 complex is in charge of the termination of short genes, and is recruited to Ser5-P CTD

Some proteins required for termination are recruited to the elongation complex by an interaction with Ser2-P CTD (Richard and Manley 2009). As Ser2-P is maximum only after 1 kb from the TSS (Jeronimo et al. 2013), short genes require an alternative way of termination (see below).

The first tyrosine in the CTD (Tyr1) is phosphorylated along the transcription unit with a pattern that closely resembles that of Ser2-P, the important difference being that it decreases before the TTS (Mayer et al. 2012). Several clues hint at the possibility that in order to be ready for termination, RNA pol II needs, along with high levels of Ser2-P, low levels of Tyr1-P (Mayer et al. 2012). This would explain why some factors of the cleavage and polyadenylation machinery are not recruited as soon as Ser2-P is at its maximum.

From this point, two alternative models of transcription termination have been proposed: allosteric and torpedo models (Richard and Manley 2009). In the allosteric model, part of the 3′ end processing machinery is recruited to the elongation complex, but only after it has transcribed the polyadenylation (poly(A)) site. Since there is no conserved termination sequence, the poly(A) site is the only feature with which we can predict the TTS. Alternatively, the elongation to termination transition can be due to elongation factors exiting after the poly(A) site. After this recruitment and/or exit step, 3′ end cleavage takes place and the elongation complex dissociates from DNA (Richard and Manley 2009). In the torpedo model, the 3′ end cleavage precedes the entry of 5′–3′ exonuclease activity that takes over the polymerase and causes its dissociation from DNA (Richard and Manley 2009). There is firm evidence to support the crucial role of both the poly(A) site and Rat1 exonuclease, which suggests that the termination process can be a combination of both models (Fig. 1.3b) (Richard and Manley 2009).

For short genes, transcription termination cannot rely on Ser2-P-dependent recruitment because they are poor in this CTD modification. Instead, the Nrd1 complex, recruited by interaction with Ser5-P (Vasiljeva et al. 2008), is essential for short mRNAs and snRNA termination (Richard and Manley 2009) (Fig. 1.3c). The mechanism of termination seems to differ since Nrd1 binds specific sequences found in the post poly(A) site RNA (Richard and Manley 2009). Consistently with the early recruitment of the Nrd1 complex due to Ser5-P peaking at early elongation, this termination mechanism is dominant, since insertion of an Nrd1 binding sequence in a long mRNA causes premature termination (Steinmetz and Brow 1998). To bind Nrd1, the CTD needs to be not only Ser5-P, but also in a *cis* conformation (Kubicek et al. 2012), which exemplifies how complicated the CTD

code can get. The requirement for the 3′ end processing machinery appears to be gene-specific within these short genes, although the reason is unclear (Richard and Manley 2009).

Transcription termination may favour reinitiation, given the physical proximity of the termination and promoter regions. This gene-looping phenomenon requires TFIIB and Ssu72 RNA pol II phosphatase, which is a component of the 3′ end processing machinery (Singh and Hampsey 2007). In fact removal of Ser7 phosphorylation from the RNA pol II CTD seems to be important for both termination and reinitiation (Zhang et al. 2012).

mRNA Transport

Some inducible genes in *Saccharomyces* relocates to the vicinity of the nuclear pore when transcription is induced during a process that is mediated by some subunits of the SAGA complex and nucleoporin Nup1 (Cabal et al. 2006). The so-called gene-gating phenomenon reflects the physical coupling between transcription and mRNA export.

In yet another layer of complexity, some processing factors may help recruiting others to the polymerase. Such is the case of a splicing factor that helps the recruitment of TREX, a complex in charge of mRNA transport to the cytoplasm (Hsin and Manley 2012) (see also Chap. 4). The main component of TREX is the THO complex (Chavez et al. 2000). The absence of any THO subunit provokes transcription elongation defects and transcription-dependent DNA instability (Chavez and Aguilera 1997) due to the formation of R-loops between nascent RNA and transcribed DNA (Huertas and Aguilera 2003). TREX cooperates with export factors Mex67 and Mtr2 in targeting mRNAs to the nuclear pore (Stewart 2010). Other mRNA export factors are also recruited co-transcriptionally. This is the case of Yra1, which binds cleavage-polyadenylation factor subunit Pcf11 and regulates 3′-end processing (Johnson et al. 2011). All together, the intricate net of interactions between elongating RNA pol II and the mRNA processing factors, and also among themselves, reflects the importance of cotranscriptionality for gene expression (Perales and Bentley 2009).

Interplay Between mRNA Synthesis and Decay

Any global transcriptional or mRNA decay variation should involve a significant change in mRNA concentrations. *Saccharomyces* genetics contradicts this prediction since transcription impairment with an RNA pol II point mutation lowers mRNA synthesis rates, but it has a very minor impact on mRNA levels (Sun et al. 2012; Goler-Baron et al. 2008). Likewise, impairment of mRNA degradation lowers decay rates, but its consequences on mRNA levels are weak (Sun et al. 2012) (see Chap. 7 for more details).

The action of a hypothetical factor that compensates transcription and decay in response to any alteration in mRNA homeostasis has been proposed to explain these results (Sun et al. 2012, 2013). Considering gene expression as a single system, in which all the steps are mechanistically coupled, offers an alternative explanation (Dahan and Choder 2013; Komili and Silver 2008) (Fig. 1.4a). The extensive network of coupling among transcription, RNA processing and mRNA export sustains this view (see above). Similarly, degradation is also coupled to other processes, such as translation (Parker and Sheth 2007).

Over the last few years, evidence has accumulated which demonstrates that *Saccharomyces* mRNA decay and mRNA biosynthesis machineries have an impact on each other. The UAS of yeast promoters and the factors binding them determine their mRNA decay kinetics (Bregman et al. 2011; Trcek et al. 2011). For instance, *Saccharomyces* transcription factor Rap1, which controls hundreds of genes, and a short *cis*-acting element comprising two Rap1-binding sites, are necessary and sufficient to induce enhanced decay of a reporter mRNA (Bregman et al. 2011). The periodic decay of *SWI5* and *CLB2* mRNAs across the cell cycle is also controlled by their promoters (Trcek et al. 2011). In this case, the control is mediated by RNA binding protein Dbf2, which seems to load onto mRNAs co-transcriptionally. These cases illustrate the hypothesis of mRNA imprinting, by which gene-specific transcription factors can modulate the loading of RNA-binding proteins onto nascent mRNA, and these, in turn, regulate mRNA decay after mRNA is exported to the cytoplasm (Choder 2011). RNA polymerase subunits Rpb4 and Rpb7 are excellent examples of mRNA imprinting factors as they are co-transcriptionally loaded onto mRNA and stimulate mRNA decay (Lotan et al. 2005, 2007; Goler-Baron et al. 2008) (Fig. 1.4b).

Saccharomyces Rpb4 and Rpb7 form a heterodimer that, apart from its role as an RNA pol II component, plays a more general role in gene expression (Choder 2004). It shuttles between the nucleus and the cytoplasm (Selitrennik et al. 2006) and can stimulate mRNA export (Farago et al. 2003), translation (Harel-Sharvit et al. 2010) and decay (Lotan et al. 2005, 2007). The interaction of Rpb4/7 with mRNA occurs only in the RNA pol II context, and is required for the Rpb4/7 stimulation of translation and mRNA decay (Harel-Sharvit et al. 2010; Goler-Baron et al. 2008). This co-transcriptional loading of Rpb4/7 onto transcripts allows RNA pol II to regulate the translation and decay rates of these mRNAs. Based on the multifunctional nature of *Saccharomyces* Rpb4/7, the existence of an mRNA coordinator complex capable of regulating the entire life of mRNAs has been proposed (Choder 2011; Perez-Ortin et al. 2013).

mRNA coordinators explain the communication from transcription to decay, and the robust homeostasis of mRNAs in response to transcriptional fluctuations. A reciprocal link exists from decay to transcription that allows the preservation of mRNA homeostasis in response to perturbations in mRNA stability. Whole-genome analyses have demonstrated that the concentrations of most *Saccharomyces* mRNAs are robust to perturbations of the main mRNA degradation pathway, as defects in various components of this machinery lead to the down-regulation of transcription (Haimovich et al. 2013). Several components of this pathway shuttle between the

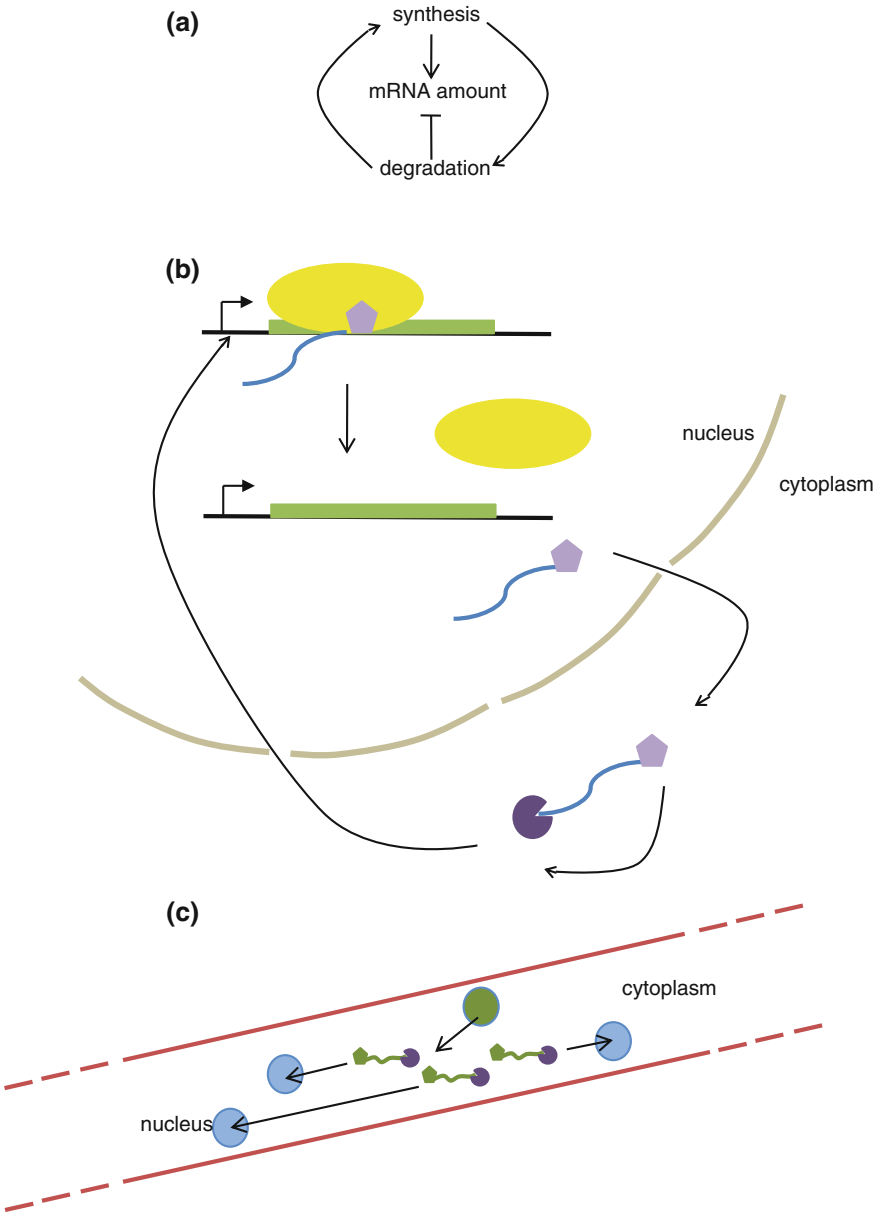


Fig. 1.4 Gene expression as a circular process. **a** Traditionally, synthesis and degradation were thought to independently influence the mRNA steady-state concentration. However, recent data indicate that synthesis factors affect degradation, and vice versa. **b** Nascent RNA can be imprinted by the binding of a transcription factor, which can later influence degradation in the cytoplasm. Some nucleases involved in cytoplasmic degradation have been shown to affect transcription. **c** A model for nucleus-to-nucleus communication in fungal syncytial hyphae, mediated by mRNA degradation

cytoplasm and the nucleus in a proper mRNA degradation-dependent manner (Haimovich et al. 2013). In the nucleus, they associate with chromatin where they stimulate transcription initiation and elongation (Haimovich et al. 2013).

The mRNA decay machinery also involves factors that degrade transcripts from the poly(A) tail. Ccr4 is the main deadenylase that catalyses this reaction in *Saccharomyces* (Tucker et al. 2002). Ccr4 belongs to Ccr4-NOT, an evolutionarily conserved complex composed of nine subunits. Based on independent experimental approaches, Ccr4-NOT has been connected to different gene transcription aspects, including initiation (Badarinarayana et al. 2000; Deluen et al. 2002) and elongation (Rodriguez-Gil et al. 2010; Denis et al. 2001; Gaillard et al. 2009). If the two roles of Ccr4-NOT turn out to be coupled, the specific role of Ccr4-NOT in transcription antagonising RNA pol II backtracking (Rodriguez-Gil et al. 2010; Kruk et al. 2011) may suggest that RNA pol II backtracking is involved in the communication between mRNA decay and transcription.

A comparison between the mRNA decay kinetics of two related *Saccharomyces* species has revealed that more than 5 % of their orthologous mRNAs have undergone a parallel evolutionary change in both mRNA decay and transcription (Dori-Bachash et al. 2011). Interestingly, the mRNA synthesis and decay capacity of Rpb4p and Ccr4p have also evolved in parallel in yeast. This suggests that the transcription-decay coupling mechanism allows greater evolutionary plasticity in gene expression levels in fungi (Dahan et al. 2011). It is, therefore, not surprising that pathogenic fungi like *Cryptococcus neoformans* require this coupling role of Rpb4 and Ccr4 to adapt to the host cell and pathogenicity (Bloom et al. 2013).

The mRNA decay-transcription interaction also opens up an interesting perspective for the coordination of multinucleate cells, a frequent situation in fungi. Imprinted RNA could be transcribed in one nucleus, degraded in the cytoplasm, and the decay factor in charge could enter a second nucleus. In this case, decay-transcription coupling would facilitate the transcriptional coordination of all the nuclei present in a syncytial cell (Fig. 1.4c).

References

Adelman K, Marr MT, Werner J, Saunders A, Ni Z, Andrulis ED et al (2005) Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS. Mol Cell 17:103–112

Albert I, Mavrich TN, Tomsho LP, Qi J, Zanton SJ, Schuster SC et al (2007) Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. Nature 446:572–576

Arimura Y, Tachiwana H, Oda T, Sato M, Kurumizaka H (2012) Structural analysis of the hexasome, lacking one histone H2A/H2B dimer from the conventional nucleosome. Biochemistry 51:3302–3309

Ast G (2004) How did alternative splicing evolve? Nat Rev Genet 5:773–782

Badarinarayana V, Chiang YC, Denis CL (2000) Functional interaction of CCR4-NOT proteins with TATAA-binding protein (TBP) and its associated factors in yeast. Genetics 155:1045–1054

Bataille AR, Jeronimo C, Jacques PE, Laramee L, Fortin ME, Forest A et al (2012) A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. Mol Cell 45:158–170

Bhaumik SR (2011) Distinct regulatory mechanisms of eukaryotic transcriptional activation by SAGA and TFIID. Biochim Biophys Acta 1809:97–108

Bintu L, Kopaczynska M, Hodges C, Lubkowska L, Kashlev M, Bustamante C (2011) The elongation rate of RNA polymerase determines the fate of transcribed nucleosomes. Nat Struct Mol Biol 18:1394–1399

Bloom AL, Solomons JT, Havel VE, Panepinto JC (2013) Uncoupling of mRNA synthesis and degradation impairs adaptation to host temperature in Cryptococcus neoformans. Mol Microbiol 89:65–83

Brahmachari SK, Sarkar PS, Raghavan S, Narayan M, Maiti AK (1997) Polypurine/polypyrimidine sequences as cis-acting transcriptional regulators. Gene 190:17–26

Brannan K, Bentley DL (2012) Control of transcriptional elongation by RNA polymerase II: a retrospective. Genet Res Int 2012:170173

Bregman A, Avraham-Kelbert M, Barkai O, Duek L, Guterman A, Choder M (2011) Promoter elements regulate cytoplasmic mRNA decay. Cell 147:1473–1483

Brill SJ, Sternglanz R (1988) Transcription-dependent DNA supercoiling in yeast DNA topoisomerase mutants. Cell 54:403–411

Cabal GG, Genovesio A, Rodriguez-Navarro S, Zimmer C, Gadal O, Lesne A et al (2006) SAGA interacting factors confine sub-diffusion of transcribed genes to the nuclear envelope. Nature 441:770–773

Chang HW, Kulaeva OI, Shaytan AK, Kibanov M, Kuznedelov K, Severinov KV et al. (2013) Analysis of the mechanism of nucleosome survival during transcription. Nucleic Acids Res 42(3):1619–1627

Chatterjee N, Sinha D, Lemma-Dechassa M, Tan S, Shogren-Knaak MA, Bartholomew B (2011) Histone H3 tail acetylation modulates ATP-dependent remodeling through multiple mechanisms. Nucleic Acids Res 39:8378–8391

Chavez S, Aguilera A (1997) The yeast HPR1 gene has a functional role in transcriptional elongation that uncovers a novel source of genome instability. Genes Dev 11:3459–3470

Chavez S, Beilharz T, Rondon AG, Erdjument-Bromage H, Tempst P, Svejstrup JQ et al (2000) A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in *Saccharomyces cerevisiae*. EMBO J 19:5824–5834

Cheung AC, Cramer P (2011) Structural basis of RNA polymerase II backtracking, arrest and reactivation. Nature 471:249–253

Cheung V, Chua G, Batada NN, Landry CR, Michnick SW, Hughes TR et al (2008) Chromatin- and transcription-related factors repress transcription from within coding regions throughout the *Saccharomyces cerevisiae* genome. PLoS Biol 6:e277

Choder M (2011) mRNA imprinting: additional level in the regulation of gene expression. Cell Logist 1:37–40

Choder M (2004) Rpb4 and Rpb7: subunits of RNA polymerase II and beyond. Trends Biochem Sci 29:674–681

Churchman LS, Weissman JS (2011) Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature 469:368–373

Core LJ, Waterfall JJ, Lis JT (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 322:1845–1848

Cramer P (2004) Structure and function of RNA polymerase II. Adv Protein Chem 67:1–42

Dahan N, Choder M (2013) The eukaryotic transcriptional machinery regulates mRNA translation and decay in the cytoplasm. Biochim Biophys Acta 1829:169–173

Dahan O, Gingold H, Pilpel Y (2011) Regulatory mechanisms and networks couple the different phases of gene expression. Trends Genet 27:316–322

Darzacq X, Shav-Tal Y, de Turris V, Brody Y, Shenoy SM, Phair RD et al (2007) In vivo dynamics of RNA polymerase II transcription. Nat Struct Mol Biol 14:796–806

Daulny A, Tansey WP (2009) Damage control: DNA repair, transcription, and the ubiquitin-proteasome system. DNA Repair (Amst) 8:444–448

Deluen C, James N, Maillet L, Molinete M, Theiler G, Lemaire M et al (2002) The Ccr4-not complex and yTAF1 (yTaf(II)130p/yTaf(II)145p) show physical and functional interactions. Mol Cell Biol 22:6735–6749

Denis CL, Chiang YC, Cui Y, Chen J (2001) Genetic evidence supports a role for the yeast CCR4-NOT complex in transcriptional elongation. Genetics 158:627–634

Dori-Bachash M, Shema E, Tirosh I (2011) Coupled evolution of transcription and mRNA degradation. PLoS Biol 9:e1001106

Engel C, Sainsbury S, Cheung AC, Kostrewa D, Cramer P (2013) RNA polymerase I structure and transcription regulation. Nature 502:650–655

Farago M, Nahari T, Hammel C, Cole CN, Choder M (2003) Rpb4p, a subunit of RNA polymerase II, mediates mRNA export during stress. Mol Biol Cell 14:2744–2755

Fazzio TG, Kooperberg C, Goldmark JP, Neal C, Basom R, Delrow J et al (2001) Widespread collaboration of Isw2 and Sin3-Rpd3 chromatin remodeling complexes in transcriptional repression. Mol Cell Biol 21:6450–6460

Fernandez-Tornero C, Moreno-Morcillo M, Rashid UJ, Taylor NM, Ruiz FM, Gruene T et al (2013) Crystal structure of the 14-subunit RNA polymerase I. Nature 502:644–649

Fish RN, Kane CM (2002) Promoting elongation with transcript cleavage stimulatory factors. Biochim Biophys Acta 1577:287–307

Gaillard H, Tous C, Botet J, Gonzalez-Aguilera C, Quintero MJ, Viladevall L et al (2009) Genome-wide analysis of factors affecting transcription elongation and DNA repair: a new role for PAF and Ccr4-not in transcription-coupled repair. PLoS Genet 5:e1000364

Gilchrist DA, Fargo DC, Adelman K (2009) Using ChIP-chip and ChIP-seq to study the regulation of gene expression: genome-wide localization studies reveal widespread regulation of transcription elongation. Methods 48:398–408

Goler-Baron V, Selitrennik M, Barkai O, Haimovich G, Lotan R, Choder M (2008) Transcription in the nucleus and mRNA decay in the cytoplasm are coupled processes. Genes Dev 22:2022–2027

Gomez-Herreros F, de Miguel-Jimenez L, Millan-Zambrano G, Penate X, Delgado-Ramos L, Munoz-Centeno MC et al (2012a) One step back before moving forward: regulation of transcription elongation by arrest and backtracking. FEBS Lett 586:2820–2825

Gomez-Herreros F, de Miguel-Jimenez L, Morillo-Huesca M, Delgado-Ramos L, Munoz-Centeno MC, and Chavez S (2012b) TFIIS is required for the balanced expression of the genes encoding ribosomal components under transcriptional stress. Nucleic Acids Res 40(14):6508–6519

Grant PA, Duggan L, Cote J, Roberts SM, Brownell JE, Candau R et al (1997) Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. Genes Dev 11:1640–1650

Grunberg S, Hahn S (2013) Structural insights into transcription initiation by RNA polymerase II. Trends Biochem Sci 38:603–611

Gu W, Reines D (1995) Identification of a decay in transcription potential that results in elongation factor dependence of RNA polymerase II. J Biol Chem 270:11238–11244

Guarente L (1988) UASs and enhancers: common mechanism of transcriptional activation in yeast and mammals. Cell 52:303–305

Haimovich G, Medina DA, Causse SZ, Garber M, Millan-Zambrano G, Barkai O et al (2013) Gene expression is circular: factors for mRNA degradation also foster mRNA synthesis. Cell 153:1000–1011

Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW et al (2004) Transcriptional regulatory code of a eukaryotic genome. Nature 431:99–104

Harel-Sharvit L, Eldad N, Haimovich G, Barkai O, Duek L, Choder M (2010) RNA polymerase II subunits link transcription and mRNA decay to translation. Cell 143:552–563

Hsin JP, Manley JL (2012) The RNA polymerase II CTD coordinates transcription and RNA processing. Genes Dev 26:2119–2137

Huertas P, Aguilera A (2003) Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. Mol Cell 12:711–721

Izban MG, Luse DS (1992) The RNA polymerase II ternary complex cleaves the nascent transcript in a 3'–5' direction in the presence of elongation factor SII. Genes Dev 6:1342–1356

Izban MG, Luse DS (1991) Transcription on nucleosomal templates by RNA polymerase II in vitro: inhibition of elongation with enhancement of sequence-specific pausing. Genes Dev 5:683–696

Jeronimo C, Bataille AR, Robert F (2013) The writers, readers, and functions of the RNA polymerase II C-Terminal Domain Code. Chem Rev 113(11):8491–8522

Johnson SA, Kim H, Erickson B, Bentley DL (2011) The export factor Yra1 modulates mRNA 3' end processing. Nat Struct Mol Biol 18:1164–1171

Joshi RS, Pina B, Roca J (2012) Topoisomerase II is required for the production of long Pol II gene transcripts in yeast. Nucleic Acids Res 40:7907–7915

Kim H, Erickson B, Luo W, Seward D, Graber JH, Pollock DD et al (2010) Gene-specific RNA polymerase II phosphorylation and the CTD code. Nat Struct Mol Biol 17:1279–1286

Kireeva ML, Hancock B, Cremona GH, Walter W, Studitsky VM, Kashlev M (2005) Nature of the nucleosomal barrier to RNA polymerase II. Mol Cell 18:97–108

Klages N, Strubin M (1995) Stimulation of RNA polymerase II transcription initiation by recruitment of TBP in vivo. Nature 374:822–823

Komili S, Silver PA (2008) Coupling and coordination in gene expression processes: a systems biology view. Nat Rev Genet 9:38–48

Kornberg RD (2007) The molecular basis of eukaryotic transcription. Proc Natl Acad Sci USA 104:12955–12961

Kruk JA, Dutta A, Fu J, Gilmour DS, Reese JC (2011) The multifunctional Ccr4-Not complex directly promotes transcription elongation. Genes Dev 25:581–593

Kubicek K, Cerna H, Holub P, Pasulka J, Hrossova D, Loehr F et al (2012) Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. Genes Dev 26:1891–1896

Kulaeva OI, Gaykalova DA, Pestov NA, Golovastov VV, Vassylyev DG, Artsimovitch I et al (2009) Mechanism of chromatin remodeling and recovery during passage of RNA polymerase II. Nat Struct Mol Biol 16:1272–1278

Kupfer DM, Drabenstot SD, Buchanan KL, Lai H, Zhu H, Dyer DW et al (2004) Introns and splicing elements of five diverse fungi. Eukaryot Cell 3:1088–1100

Kuryan BG, Kim J, Tran NN, Lombardo SR, Venkatesh S, Workman JL et al (2012) Histone density is maintained during transcription mediated by the chromatin remodeler RSC and histone chaperone NAP1 in vitro. Proc Natl Acad Sci USA 109:1931–1936

Liu P, Kenney JM, Stiller JW, Greenleaf AL (2010) Genetic organization, length conservation, and evolution of RNA polymerase II carboxyl-terminal domain. Mol Biol Evol 27:2628–2641

Lorch Y, Maier-Davis B, Kornberg RD (2006) Chromatin remodeling by nucleosome disassembly in vitro. Proc Natl Acad Sci USA 103:3090–3093

Lotan R, Bar-On VG, Harel-Sharvit L, Duek L, Melamed D, Choder M (2005) The RNA polymerase II subunit Rpb4p mediates decay of a specific class of mRNAs. Genes Dev 19:3004–3016

Lotan R, Goler-Baron V, Duek L, Haimovich G, Choder M (2007) The Rpb7p subunit of yeast RNA polymerase II plays roles in the two major cytoplasmic mRNA decay mechanisms. J Cell Biol 178:1133–1143

Luse DS, Spangler LC, Ujvari A (2011) Efficient and rapid nucleosome traversal by RNA polymerase II depends on a combination of transcript elongation factors. J Biol Chem 286:6040–6048

Margaritis T, Holstege FC (2008) Poised RNA polymerase II gives pause for thought. Cell 133:581–584

Mason PB, Struhl K (2005) Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. Mol Cell 17:831–840

Mayer A, Heidemann M, Lidschreiber M, Schreieck A, Sun M, Hintermair C et al (2012) CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. Science 336:1723–1725

Millan-Zambrano G, Rodriguez-Gil A, Penate X, de Miguel-Jimenez L, Morillo-Huesca M, Krogan N et al (2013) The prefoldin complex regulates chromatin dynamics during transcription elongation. PLoS Genet 9:e1003776

Mitra D, Parnell EJ, Landon JW, Yu Y, Stillman DJ (2006) SWI/SNF binding to the HO promoter requires histone acetylation and stimulates TATA-binding protein recruitment. Mol Cell Biol 26:4095–4110

Morris DP, Greenleaf AL (2000) The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. J Biol Chem 275:39935–39943

Papai G, Weil PA, Schultz P (2011) New insights into the function of transcription factor TFIID from recent structural studies. Curr Opin Genet Dev 21:219–224

Parker R, Sheth U (2007) P bodies and the control of mRNA translation and degradation. Mol Cell 25:635–646

Pelechano V, Jimeno-Gonzalez S, Rodriguez-Gil A, Garcia-Martinez J, Perez-Ortin JE, Chavez S (2009) Regulon-specific control of transcription elongation across the yeast genome. PLoS Genet 5:e1000614

Pelechano V, Wei W, Steinmetz LM (2013) Extensive transcriptional heterogeneity revealed by isoform profiling. Nature 497:127–131

Perales R, Bentley D (2009) "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. Mol Cell 36:178–191

Perez-Ortin JE, Alepuz P, Chavez S, Choder M (2013) Eukaryotic mRNA decay: methodologies, pathways, and links to other stages of gene expression. J Mol Biol 425:3750–3775

Perez-Ortin JE, de Miguel-Jimenez L, Chavez S (2011) Genome-wide studies of mRNA synthesis and degradation in eukaryotes. Biochim Biophys Acta 1819:604

Phatnani HP, Jones JC, Greenleaf AL (2004) Expanding the functional repertoire of CTD kinase I and RNA polymerase II: novel phosphoCTD-associating proteins in the yeast proteome. Biochemistry 43:15702–15719

Powell W, Reines D (1996) Mutations in the second largest subunit of RNA polymerase II cause 6-azauracil sensitivity in yeast and increased transcriptional arrest in vitro. J Biol Chem 271:6866–6873

Pugh BF (2000) Control of gene expression through regulation of the TATA-binding protein. Gene 255:1–14

Rhee HS, Pugh BF (2012) Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature 483:295–301

Richard P, Manley JL (2009) Transcription termination by nuclear RNA polymerases. Genes Dev 23:1247–1269

Rodriguez CR, Cho EJ, Keogh MC, Moore CL, Greenleaf AL, Buratowski S (2000) Kin28, the TFIIH-associated carboxy-terminal domain kinase, facilitates the recruitment of mRNA processing machinery to RNA polymerase II. Mol Cell Biol 20:104–112

Rodriguez-Gil A, Garcia-Martinez J, Pelechano V, Munoz-Centeno Mde L, Geli V, Perez-Ortin JE et al (2010) The distribution of active RNA polymerase II along the transcribed region is gene-specific and controlled by elongation factors. Nucleic Acids Res 38:4651–4664

Rodriguez-Navarro S (2009) Insights into SAGA function during gene expression. EMBO Rep 10:843–850

Schwabish MA, Struhl K (2006) Asf1 mediates histone eviction and deposition during elongation by RNA polymerase II. Mol Cell 22:415–422
Schwabish MA, Struhl K (2004) Evidence for eviction and rapid deposition of histones upon transcriptional elongation by RNA polymerase II. Mol Cell Biol 24:10111–10117
Selitrennik M, Duek L, Lotan R, Choder M (2006) Nucleocytoplasmic shuttling of the Rpb4p and Rpb7p subunits of *Saccharomyces cerevisiae* RNA polymerase II by two pathways. Eukaryot Cell 5:2092–2103
Sigurdsson S, Dirac-Svejstrup AB, Svejstrup JQ (2010) Evidence that transcript cleavage is essential for RNA polymerase II transcription and cell viability. Mol Cell 38:202–210
Singh BN, Hampsey M (2007) A transcription-independent role for TFIIB in gene looping. Mol Cell 27:806–816
Sogaard TM, Svejstrup JQ (2007) Hyperphosphorylation of the C-terminal repeat domain of RNA polymerase II facilitates dissociation of its complex with mediator. J Biol Chem 282:14113–14120
Somesh BP, Reid J, Liu WF, Sogaard TM, Erdjument-Bromage H, Tempst P et al (2005) Multiple mechanisms confining RNA polymerase II ubiquitylation to polymerases undergoing transcriptional arrest. Cell 121:913–923
Steinmetz EJ, Brow DA (1998) Control of pre-mRNA accumulation by the essential yeast protein Nrd1 requires high-affinity transcript binding and a domain implicated in RNA polymerase II association. Proc Natl Acad Sci USA 95:6699–6704
Stewart M (2010) Nuclear export of mRNA. Trends Biochem Sci 35:609–617
Struhl K, Segal E (2013) Determinants of nucleosome positioning. Nat Struct Mol Biol 20:267–273
Sun M, Schwalb B, Pirkl N, Maier KC, Schenk A, Failmezger H et al (2013) Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. Mol Cell 52:52–62
Sun M, Schwalb B, Schulz D, Pirkl N, Etzold S, Lariviere L et al (2012) Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. Genome Res 22:1350–1359
Takahata S, Yu Y, Stillman DJ (2009) The E2F functional analogue SBF recruits the Rpd3(L) HDAC, via Whi5 and Stb1, and the FACT chromatin reorganizer, to yeast G1 cyclin promoters. EMBO J 28:3378–3389
Thomas MC, Chiang CM (2006) The general transcription machinery and general cofactors. Crit Rev Biochem Mol Biol 41:105–178
Tjian R (1978) The binding site on SV40 DNA for a T antigen-related protein. Cell 13:165–179
Trcek T, Larson DR, Moldon A, Query CC, Singer RH (2011) Single-molecule mRNA decay measurements reveal promoter- regulated mRNA stability in yeast. Cell 147:1484–1497
Tsankov A, Yanagisawa Y, Rhind N, Regev A, Rando OJ (2011) Evolutionary divergence of intrinsic and trans-regulated nucleosome positioning sequences reveals plastic rules for chromatin organization. Genome Res 21:1851–1862
Tucker M, Staples RR, Valencia-Sanchez MA, Muhlrad D, Parker R (2002) Ccr4p is the catalytic subunit of a Ccr4p/Pop2p/Notp mRNA deadenylase complex in *Saccharomyces cerevisiae*. EMBO J 21:1427–1436
Vannini A, Cramer P (2012) Conservation between the RNA polymerase I, II, and III transcription initiation machineries. Mol Cell 45:439–446
Vasiljeva L, Kim M, Mutschler H, Buratowski S, Meinhart A (2008) The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. Nat Struct Mol Biol 15:795–804
Venters BJ, Pugh BF (2009a) A canonical promoter organization of the transcription machinery and its regulators in the Saccharomyces genome. Genome Res 19:360–371
Venters BJ, Pugh BF (2009b) How eukaryotic genes are transcribed. Crit Rev Biochem Mol Biol 44:117–141

Weiner A, Hughes A, Yassour M, Rando OJ, Friedman N (2010) High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. Genome Res 20:90–100

Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, Altschuler SJ et al (2005) Genome-scale identification of nucleosome positions in *S. cerevisiae*. Science 309:626–630

Zhang DW, Mosley AL, Ramisetty SR, Rodriguez-Molina JB, Washburn MP, Ansari AZ (2012) Ssu72 phosphatase-dependent erasure of phospho-Ser7 marks on the RNA polymerase II C-terminal domain is essential for viability and transcription termination. J Biol Chem 287:8541–8551

Zhang Z, Pugh BF (2011) High-resolution genome-wide mapping of the primary structure of chromatin. Cell 144:175–186

Chapter 2
Pre-mRNA Splicing and the Spliceosome: Assembly, Catalysis, and Fidelity

Elizabeth A. Dunn and Stephen D. Rader

Abstract At the level of gene architecture, the widespread presence of interrupting sequences in eukaryotic genes serves as a defining difference between eukaryotic organisms and other domains of life. These interrupting sequences, known as introns, must be precisely removed from pre-messenger RNA (pre-mRNA) transcripts. Concomitantly, the coding regions, or exons, are joined together through a nuclear-localized process known as pre-mRNA splicing. A number of splicing factors, both protein and RNA, assemble into a multimegadalton splicing machine known as the spliceosome, which is responsible for identifying the intronic regions and positioning the pre-mRNA substrate in a favorable orientation for the splicing reactions to occur. While the chemical steps of splicing—two sequential transesterification reactions—are identical in all eukaryotes, the gene architecture and splicing apparatus can differ substantially. Here, we review our current understanding of the splicing process with an emphasis on the model organism *Saccharomyces cerevisiae*. We discuss the key features of introns, along with mechanistic aspects of the splicing cycle, namely spliceosome assembly, catalysis, and spliceosome disassembly. We also highlight recent discoveries supporting the role of kinetic proofreading in ensuring the fidelity of splicing.

S. D. Rader (✉)
Department of Chemistry, University of Northern British Columbia,
Prince George, BC, Canada
e-mail: rader@unbc.ca

E. A. Dunn
Department of Biochemistry and Molecular Biology,
University of British Columbia, Vancouver, BC, Canada

A. Sesma and T. von der Haar (eds.), *Fungal RNA Biology*,
DOI: 10.1007/978-3-319-05687-6_2,

Contents

Gene Architecture

The eukaryotic gene is composed of protein coding sequences, known as exons, which are interrupted by non-protein coding sequences called introns (Fig. 2.1a). Over the last decade and a half, a number of complete fungal genomes have been sequenced, allowing in-depth comparative analyses of intronic features, intron abundance and position, and intron evolution. These analyses have revealed that intron-containing genes are prevalent across fungal species to varying extents, although an intron-poor genome appears to be a common feature in budding yeasts (Neuvéglise et al. 2011). In most of these species, including the well-established model organism *Saccharomyces cerevisiae*, fewer than 10 % of genes contain an intron (Spingola et al. 1999; Bon et al. 2003; Neuvéglise et al. 2011). For example, only 5 % of *S. cerevisiae*'s genes contain an intron, in stark contrast to the intron-rich fission yeast, *Schizosaccharomyces pombe*, in which 43 % of genes contain introns (Wood et al. 2002). *S. pombe* most closely resembles metazoans in this respect. For example, approximately 95 % of human genes contain introns (Venter et al. 2001). Much like mammalian genes, composed of an average of seven introns per gene, *S. pombe* genes tend to contain multiple introns, with 34 genes containing between 7 and 15 introns (Sakharkar et al. 2004; Wood et al. 2002). In fact, only 45 % of intron-containing genes in *S. pombe* contain a single intron while 49 % contain two, three, or four introns (Wood et al. 2002).

In contrast, hemiascomycetous yeast species (including *Saccharomyces* and *Candida* species) possess very few genes with a single intron and no gene with more than two (Spingola et al. 1999; Neuvéglise et al. 2011). Intriguingly, the introns of all budding yeasts show a positional bias toward the 5′ end of the gene (Spingola et al. 1999; Neuvéglise et al. 2011). In *S. pombe*, however, introns in genes that contain seven introns or more are distributed throughout the gene with no positional bias, much like metazoans. In contrast, those genes that contain

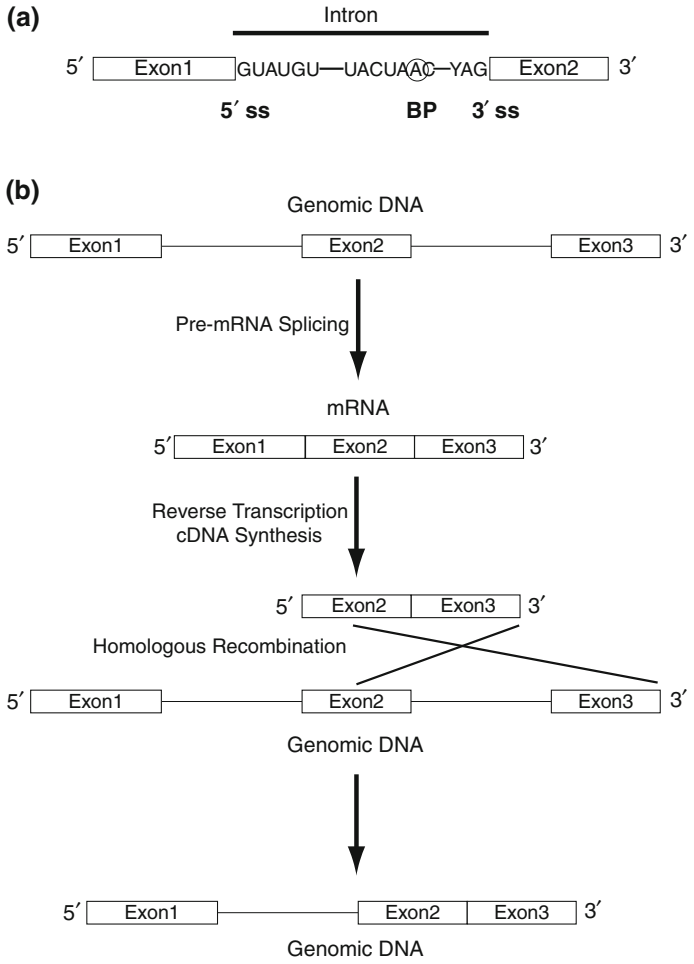


Fig. 2.1 **a** Eukaryotic gene structure, showing introns (*top*) and exons (*boxed*). The 5′ splice site (5′ss), branchpoint sequence (BP, reactive adenosine *circled*), and 3′ splice site (3′ss) are indicated along with the consensus sequence at each of these sites. **b** Homologous recombination model of intron loss. Exons are *boxed*, while *horizontal lines* connecting the exons indicate introns

fewer than seven introns show the same 5′ positional bias observed in budding yeasts (Wood et al. 2002). Fink (1987) attempted to explain this phenomenon through a homologous recombination model that suggested that reverse transcribed cDNAs generated from an mRNA template undergo homologous recombination with the yeast genome at the site of the gene that produced the original mRNA (Fig 2.1b). Since reverse transcriptase often falls off the template before reaching the 5′ end of long mRNAs, Fink (1987) argued that an intronless 3′ end of the gene would be over-represented in the cDNA population, and therefore the

recombination event would be more likely to replace the intron-containing 3′ end of the gene with an intron-less one. Consequently, a higher frequency of 5′ introns would be retained in the original gene.

The homologous recombination model makes three major predictions: retained introns show a 5′ positional bias, intron loss is precise, and adjacent introns are lost simultaneously (Fink 1987). All of these predictions have been substantiated in a number of fungal species, as well as in vertebrates (Coulombe-Huntington and Majewski 2007; Zhu and Niu 2013). Even in intron-rich fission yeasts, 656 out of 660 identified intron losses were precise and located in the 3′ portion of the gene, and 38 different losses of adjacent introns were reported (Zhu and Niu 2013). According to the homologous recombination model, one would expect that more frequently transcribed genes would lose their introns first, since there would be more of these mRNAs present to serve as templates for reverse transcription. However, this is not the case with the ribosomal protein genes that represent about 90 % of all mRNAs derived from intron-containing genes in *S. cerevisiae* (Ares et al. 1999). Among intron-containing genes in budding yeasts, ribosomal protein genes are over-represented, reaching as a high as 41 and 61 % in *S. cerevisiae* and *S. servazzii*, respectively (Bon et al. 2003). The introns of these highly expressed ribosomal protein genes have been retained through selective pressure and they appear to play a functional role in positive regulation of ribosome biogenesis (Vinogradov 2001).

While the homologous recombination model of mass intron loss in many fungal species still dominates today, several alternative models have been put forward in recent years. Under the genomic deletion model, introns are lost individually through the unequal exchange of alleles, often resulting in imprecise intron loss (Soochin et al. 2004). Evidence for such a mechanism has been found in only a small number of multicellular organisms and in only 4 of more than 600 cases of intron loss across 4 fission yeast species (Zhu and Niu 2013 and therein). In a third model, introns can be lost either precisely or imprecisely through nonhomologous end joining repair of double strand breaks (Farlow et al. 2011). While evidence for this model has been presented for only a few metazoan species, there are no supporting examples across fission yeast species (Farlow et al. 2011; Zhu and Niu 2013). It should be noted that intron gain is also common, and, in fungi, the presence of nearly identical introns in up to 500 copies per cell supports a mechanism of intron gain that involves duplication of intron-like elements (Collemare et al. 2013).

Unlike their mammalian counterparts, introns found in the budding yeasts are typically very short, showing a bimodal distribution of intron length with averages of 50–100 nucleotides and 250–400 nucleotides (Spingola et al. 1999; Bon et al. 2003; Neuvéglise et al. 2011). Intriguingly, these two classes of intron length can be assigned to two specific groups of genes: those encoding ribosomal proteins (longer introns) and those encoding nonribosomal proteins (shorter introns) (Spingola et al. 1999). Short introns are common to all known budding yeast, although exceptions have been reported (Bon et al. 2003; Neuvéglise et al. 2011). For example, introns in *Y. lipolytica* range in size from 41 to 3478 nucleotides,

with 16 introns larger than 1 kb (Mekouar et al. 2010). Even more distant fungal relatives, such as *S. pombe*, have short introns with an average length of 50 nucleotides, although the range in size, 29–819 nucleotides, is much smaller than in *Y. lipolytica* (Deutsch and Long 1999; Wood et al. 2002).

Spliceosomal introns have three defining features: the 5′ splice site (5′ss), the branch point (BP), and the 3′ splice site (3′ss) (Fig. 2.1a). The sequences at these sites are generally well conserved from budding yeast to humans, where in almost all cases the intron begins with a GT dinucleotide and ends with an AG dinucleotide (Spingola et al. 1999; Burset et al. 2000). Exceptions to this so-called GT–AG rule can be found both within and across species; for example, five *S. cerevisiae* and three *S. pombe* introns begin with the dinucleotide GC (Spingola et al. 1999; Wood et al. 2002). In addition, the sequence context at the splice sites and branch point is very important in *S. cerevisiae* and other closely related budding yeast, for which the consensus sequences GTATGT, TACTAAC (where A is the branch nucleotide), and YAG (where Y is a pyrimidine), at the 5′ss, BP, and 3′ss, respectively, are adhered to very closely (Bon et al. 2003; Neuvéglise et al. 2011). Some flexibility is observed at the 5′ss and BP in the more distantly related species *D. hansenii*, *P. angusta*, and *Y. lipolytica*, however, the 3′ss sequence in these species conforms to the YAG consensus (Bon et al. 2003). Notably, the sequences GTAAGT and GTGAGT at the 5′ss are the dominant sequences in *P. angusta* and *Y. lipolytica*, respectively, where these sequences represent 69 and 75 % of the introns in these genomes (Bon et al. 2003).

Pre-mRNA Splicing and the Spliceosome

Eukaryotic pre-mRNA splicing is catalyzed by the spliceosome, a large complex of five small nuclear RNAs (snRNAs U1, U2, U4, U5, and U6) and more than 100 core proteins (Jurica and Moore 2003). Many of these proteins associate specifically with an snRNA to form small nuclear ribonucleoprotein (snRNP) particles, while others associate with the spliceosome in an snRNA-independent manner that is often mediated through protein–protein interactions. Assembly of the spliceosome on a newly transcribed pre-mRNA substrate requires the addition of four major splicing subcomplexes: U1 snRNP, U2 snRNP, the pre-formed U4/U6•U5 triple snRNP, and the Prp19 Complex (Cheng and Abelson 1987; Hoskins et al. 2011). Through numerous structural rearrangements, the spliceosome interacts dynamically with the transcript, recognizing the 5′ss, BP, and 3′ss, and positioning the pre-mRNA substrate in a favorable orientation for the splicing reactions to proceed (Brody and Abelson 1985; Grabowski et al. 1985).

Recent real-time kinetic analyses of spliceosome assembly using multiwavelength fluorescence microscopy support a long held view that spliceosome assembly occurs through the highly ordered association of subcomplexes with the transcript (Cheng and Abelson 1987; Hoskins et al. 2011). These studies show that commitment of the transcript to splicing increases as assembly progresses, and that

association of the subcomplexes with the transcript is reversible (Hoskins et al. 2011). Notably, although higher order splicing complexes such as a penta-snRNP have been purified and characterized (Stevens et al. 2002), Crawford et al. (2013) find no evidence for the association of preformed complexes with the pre-mRNA. Thus, such higher-order complexes probably represent a stable association of the constituents already assembled on pre-mRNA substrates. Regardless of whether a *bona fide* penta-snRNP exists independent of substrate in vivo, such a species would presumably have to undergo the same conformational and compositional rearrangements outlined in the stepwise assembly model in order to ensure that all proofreading stages are passed so that high fidelity splicing can be achieved. The details of spliceosome assembly, activation, catalysis, and disassembly will be discussed here, focusing on the findings in the model organism *S. cerevisiae*. Consequently, the nomenclature used will be that of *S. cerevisiae* unless otherwise stated.

Association of U1 snRNP with the pre-mRNA Transcript

Assembly of the spliceosome begins with the association of U1 snRNP with the pre-mRNA transcript through a base-pairing interaction between the 5′ end of U1 snRNA and the 5′ss of the transcript (Fig. 2.2; Siliciano and Guthrie 1988; Crawford et al. 2013). This association proceeds in the absence of ATP hydrolysis, and is dependent on the presence of an intact 5′ss that maintains base pairing at intron positions one and five, but not at position four (Fig. 2.2; Siliciano and Guthrie 1988; Crawford et al. 2013). Notably, even though the 5′ss consensus sequence in mammals is far more degenerate than in yeast, the first ten nucleotides of U1 snRNA are invariant across eukaryotes (Guthrie and Patterson 1988). In mammals, the site of cleavage at the 5′ss is determined by complementarity to U1 rather than by the intron sequence, with specific cleavage occurring opposite the C8–C9 nucleotides of U1 (Weber and Aebi 1988). This is not the case for *S. cerevisiae*, where authentic 5′ss cleavage appears to require a G at position five of the intron, along with the U1 snRNP specific proteins Nam8 and Luc7, which stabilize the 5′ss/U1 interaction through contacts with the intron and 5′ exon, respectively (Siliciano and Guthrie 1988; Puig et al. 1999, 2007).

U1 snRNA is fairly well conserved across eukaryotes, consisting of an almost invariant short single-stranded 5′ end, three stem loop structures (stems I, II, and III) that are closed by a long-range interaction, a single-stranded region containing the Sm protein binding site, and a terminal stem loop (stem IV) (Fig. 2.2; Guthrie and Patterson 1988). Stem III is highly divergent in the hemiascomycetous yeasts, ranging from a short stem of 14 nucleotides in *Y. lipolytica* to a long unbranched stem of 104 nucleotides in the *Candida* species, and a long multibranched stem loop in *S. cerevisiae* (Mitrovich and Guthrie 2007). Intriguingly, this large insertion, referred to as the U1 snRNA fungal domain (Guthrie and Patterson 1988), is accompanied by the presence of several yeast-specific U1 snRNP proteins. Prp42,

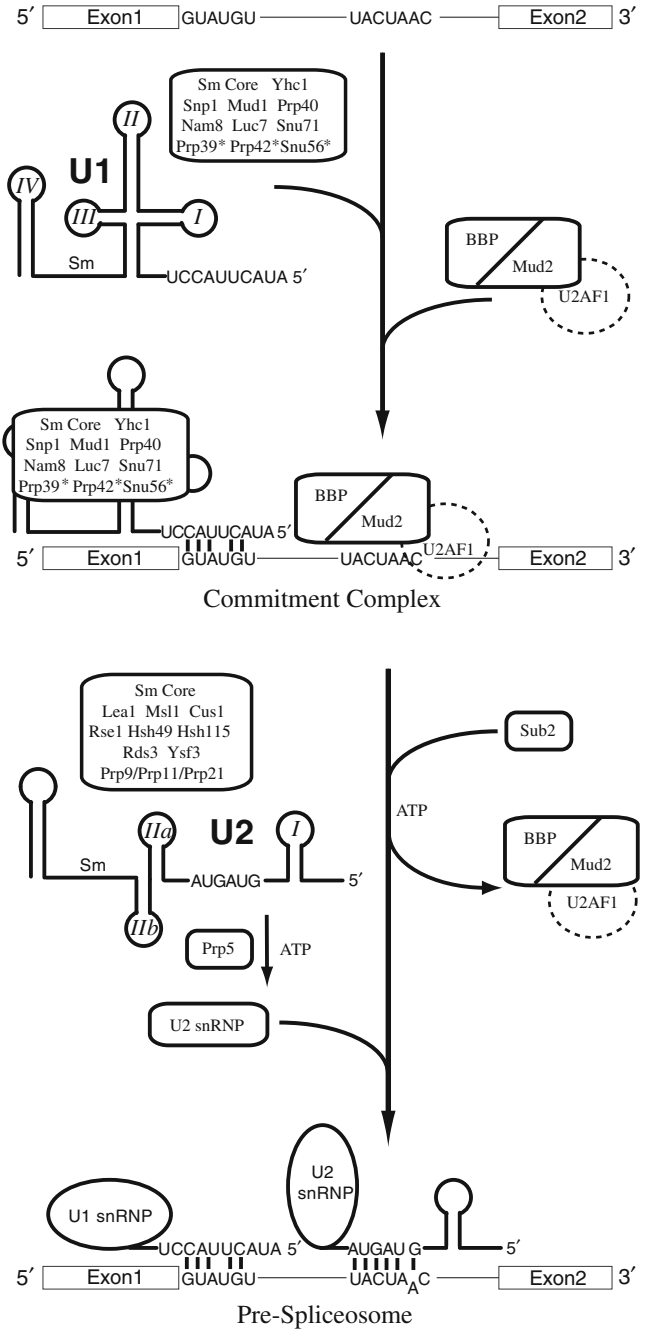


Fig. 2.2 Pre-spliceosome assembly showing ATP-independent addition of U1 snRNP and the BBP/Mud2 dimer to the 5′ss and BP, respectively (*top*), followed by the ATP-dependent addition of U2 snRNP to the BP (*bottom*). Proteins are indicated by *large rectangles*; those unique to *S. cerevisiae* are indicated with an *asterisk*, while those that are absent are indicated with a *dashed oval*. Base pairing interactions between RNA nucleotides are indicated by *vertical lines*

which is thought to have arisen as a duplication of the yeast-specific protein Prp39 in a common ancestor of *S. cerevisiae* and *C. albicans*, might interact with the extended stem III, since a homolog in *Y. lipolytica* does not exist (Mitrovich and Guthrie 2007). Stem III of *Y. lipolytica* is more similar in size to most other eukaryotes that also lack a Prp42 homolog (Fabrizio et al. 2009). Likewise, the *S. cerevisiae* specific protein Snu56 might associate with U1 snRNA through its extended and branched Stem III (Mitrovich and Guthrie 2007).

Of the ten U1 snRNP specific proteins identified in *S. cerevisiae*, seven have mammalian homologs, although not all of the mammalian homologs associate specifically with U1 snRNP, and the mode of interaction between the protein and U1 snRNA has not necessarily been conserved (Mitrovich and Guthrie 2007; Fabrizio et al. 2009). For example, the human homolog of Mud1, U1A, interacts with human U1 snRNA through direct contacts between the amino-terminal RNA recognition motif (RRM) and the loop nucleotides of U1 stem loop II (Oubridge et al. 1994). While this binding interaction appears to be conserved in *C. albicans*, the nucleotide sequence has become quite degenerate in *S. cerevisiae*, accompanied by an insertion of 30 amino acids and degeneration of the surrounding amino acid sequence in the Mud1 RRM, suggesting that the mode of interaction between U1 snRNA and Mud1 in *S. cerevisiae* is different from in humans (Mitrovich and Guthrie 2007). Interestingly, the opposite situation is observed for the human protein homolog of Snp1, U1-70 K, which binds the loop residues of stem loop I in humans (Surowy et al. 1989). In this case, the *S. cerevisiae* interaction is invariant while *C. albicans* shows some sequence degeneration at the site of interaction (Mitrovich and Guthrie 2007).

In addition to recognizing the 5′ss of splicing substrates, U1 snRNP has been proposed to play a role in increasing splicing fidelity. Single molecule fluorescence resonance energy transfer (smFRET) experiments have revealed that the 5′ss and BP are held apart upon U1 snRNP association with the transcript, as demonstrated by a reduction in FRET efficiency upon U1 snRNP binding (Crawford et al. 2013). Furthermore, these sites remain separated during spliceosome assembly up to the point of spliceosome activation (Crawford et al. 2013). These authors propose that this additional role for U1 snRNP—to physically separate chemically reactive groups—is crucial to ensuring that splicing cannot occur until the spliceosome has assembled correctly. Indeed the U5 snRNP protein Prp28 has been shown to play a role in proofreading at the 5′ss, an event that would necessarily occur later in spliceosome assembly, i.e., once the triple-snRNP has assembled onto the splicing substrate, but before spliceosome activation (Yang et al. 2013).

Recognition of the Branchpoint

Prior to recruitment of the other major splicing complexes to the pre-mRNA substrate, the BBP-Mud2 heterodimer binds the BP sequence in an ATP independent manner, making direct contacts with the pre-mRNA in this region

(Fig. 2.2; Abovich et al. 1994; Wang et al. 2008). In most species, including most fungi, the homologous BBP-Mud2 complex contains a third protein, U2AF1, which interacts with the AG dinucleotide located at the 3′ss (Wu et al. 1999). Notably, U2AF1 is not present in *S. cerevisiae*. U2AF1 is highly conserved, when present, with the *S. pombe* and human proteins showing 75 % similarity (Käufer and Potashkin 2000). The presence of U2AF1 appears to correlate with a short distance between the BP and 3′ss (less than 15 nucleotides), suggesting that this heterotrimer is responsible for identifying both the BP and 3′ss (Neuvéglise et al. 2011). In species such as *S. cerevisiae*, where this distance is much longer (on average 30 nucleotides) and contains a conserved tract of polypyrimidines (PPT) near the 3′ss, association of the BBP-Mud2 complex is 3′ss independent, and interactions between the BBP-Mud2 complex and the BP and PPT appear to be stronger (Rymond and Rosbash 1985; Neuvéglise et al. 2011).

Commitment of a splicing substrate to the splicing pathway requires the stable association of U1 snRNP and the BBP–Mud2 complex at the 5′ss and BP regions, respectively, although commitment complex formation is reversible (Legrain et al. 1985; Crawford et al. 2013). One of the key features of the commitment complex is the formation of a bridge connecting the 5′ss and BP through protein-protein contacts that involve a direct physical interaction between BBP and the U1 snRNP specific protein Prp40 (Abovich and Rosbash 1997; Schwer et al. 2013). The presence of homologs of these bridging proteins, Prp40, BBP, and Mud2 in *S. pombe* and humans suggest that the cross-talk between the 5′ and 3′ regions of the intron is important at very early stages of intron recognition and spliceosome assembly across eukaryotes (Käufer and Potashkin 2000). Once this network of contacts has been established, the assembling spliceosome is then ready to accept the U2 snRNP complex.

Stable association of U2 snRNP with the pre-mRNA to form the pre-spliceosome is the first ATP-dependent step in spliceosome assembly (Fig. 2.2; Crawford et al. 2013). Two different ATPases, Sub2 and Prp5, are required at this stage to allow direct base-pairing interactions between U2 snRNA and the intron to form (Parker et al. 1987; Kistler and Guthrie 2001; O'Day et al. 1996). Sub2 is thought to function in the removal of Mud2 and BBP from the pre-mRNA, exposing the BP region of the transcript, while Prp5 appears to play a role in U2 snRNA remodeling to make the U2 snRNA BP-binding sequence more accessible. Interestingly, association of the Prp9/Prp11/Prp21 complex (SF3a complex in humans) with U2 snRNA is required for Prp5 activity, and RNase H treatment of U2 snRNA has shown that prior to assembly onto the transcript, three different Prp9/Prp11/Prp21-dependent and Prp5-dependent U2 conformations exist (Wiest et al. 1996). Association of the Prp9/Prp11/Prp21 complex with U2 converts a more open U2 snRNA that is unable to form pre-spliceosomes into a more closed particle that becomes the Prp5 substrate (Wiest et al. 1996). Association of Prp5 and subsequent ATP hydrolysis then converts U2 snRNP into a second more open conformation that is competent for association with the BP of the pre-mRNA (Wiest et al. 1996).

U2 snRNA is composed of four stem loop structures, the first three of which are separated by short single-stranded regions (Fig. 2.2; Guthrie and Patterson 1988). Like U1 snRNA, U2 snRNA is highly conserved across eukaryotes, except in *S. cerevisiae* where a large insertion of approximately 1 kb, referred to as the U2 fungal domain, replaces the third stem loop (Guthrie and Patterson 1988). Surprisingly, the entire fungal domain can be deleted without affecting growth in yeast, and a yeast U2 snRNA deletion can be complemented with human U2 snRNA (Shuster and Guthrie 1988, 1990). In contrast to U1 snRNP, the U2 snRNP components are much more highly conserved, with eleven proteins specifically associating with U2 snRNA in both yeast and humans in addition to the Sm protein core (Fabrizio et al. 2009). Stable association of U2 snRNP with the pre-mRNA involves direct U2 snRNP protein contacts with the pre-mRNA upstream of the BP in addition to base pairing (Gozani et al. 1996, 1998).

Assembly of the U4/U6•U5 Triple snRNP

In contrast to U5 and U6 snRNPs, which can exist as free particles, U4 is almost always found in association with U6 in either the U4/U6 di-snRNP or in the U4/U6•U5 tri-snRNP (Fig. 2.3; Cheng and Abelson 1987; Fortner et al. 1994). This association involves an extensive base pairing interaction that spans U6 nucleotides 55–80 and U4 nucleotides 1–17 and 57–64, generating U4/U6 stem I and stem II (Fig. 2.3; Brow and Guthrie 1988). The U4/U6 di-snRNP protein complement is small, comprised only of the U6-associated LSm proteins, the U4-associated Sm proteins, and four proteins found in the di-snRNP that are not part of free U6 snRNP (Stevens et al. 2001; Fabrizio et al. 2009). It is not clear whether these four proteins, Prp31, Prp3, Prp4, and Snu13, associate with U4 snRNA in a free U4 snRNP particle since isolation and characterization of free U4 snRNP has not been possible due to its very low abundance. Alternatively, these proteins might recognize and bind the U4/U6 duplex at some point during U4/U6 di-snRNP formation.

The U6 snRNP-specific protein Prp24, consisting largely of four RRMs, is the only other protein found associated with U6 snRNA in free U6 snRNP, aside from the Lsm2-8 protein complex that binds the 3′ uridine-rich tail of U6 (Fig. 2.3; Stevens et al. 2001; Mayes et al. 1999). While Prp24 does not stably associate with the U4/U6 di-snRNP, the rate of base-pair formation between U4 and U6 snRNAs is greatly enhanced in its presence (Shannon and Guthrie 1991; Raghunathan and Guthrie 1998). It is not yet understood how Prp24 facilitates this interaction, but it is known to do so in an ATP-independent manner (Raghunathan and Guthrie 1998). Surprisingly, the structure of yeast Prp24, which consists of four RRMs, is strikingly different from the mammalian homolog, SART3, which is three times as large and consists of only two RRMs and a long amino-terminal extension not present in *S. cerevisiae* (Bell et al. 2002; Rader and Guthrie 2002). The first two RRMs of Prp24 most closely resemble the SART3 RRMs, and in both yeast and

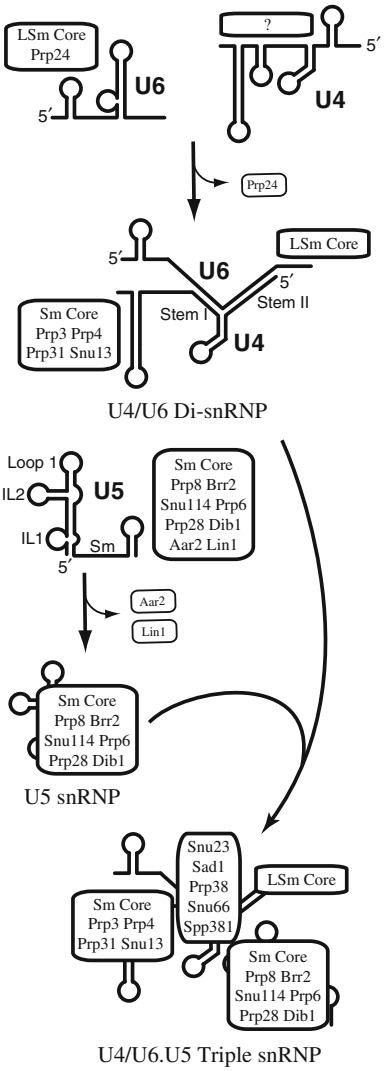


Fig. 2.3 Tri-snRNP assembly showing snRNA secondary structures (U6 and U4, *top*, U5, *middle*, stems and loops labeled as in text), and associated proteins in *large rectangles*

humans have been shown to bind U6 snRNA with high affinity (Bell et al. 2002; Kwan and Brow 2005). The *S. pombe* homolog is similar in size to SART3, consisting of a large amino-terminal extension in addition to the four RRMs that are common among most other fungi (Rader and Guthrie 2002). Whether the additional RRMs in yeast Prp24 perform a similar function to the amino-terminal extension found in other homologs remains to be determined.

Unlike U1 and U2 snRNAs, neither fungal U4 nor U6 snRNAs deviate in size relative to other eukaryotes (Guthrie and Patterson 1988). Essentially all of the very little size variation in U6 snRNA is found in the 5′ stem loop where the length

of the stem can vary by several base pairs (Brow and Guthrie 1988). In addition to the size conservation, U6 snRNA exhibits a striking level of primary sequence conservation with close to 80 % sequence identity across the middle third of the RNA across eukaryotes (Brow and Guthrie 1988). This region of U6 engages in base pairing interactions with U4, and consequently it is not surprising that the corresponding region of U4 snRNA is highly conserved in primary sequence as well. Outside of this region, however, the primary sequence is quite degenerate (Guthrie and Patterson 1988). On the U4 side of the U4/U6 duplex, stems I and II are interrupted by a stem loop structure, the 5′ stem loop, which has been absolutely conserved in structure from yeast to humans even though the nucleotide sequence in the stem differs at almost every position (Fig. 2.3; Guthrie and Patterson 1988). This high level of phylogenetic co-variation argues for an important function for this structure, which has been shown to bind the protein Snu13 (Vidovic et al. 2000). U4 snRNA also contains a 3′ stem loop that varies substantially across eukaryotes, followed by the Sm protein-binding site and, in most eukaryotes excluding *S. cerevisiae*, a final stem loop structure (Guthrie and Patterson 1988).

In order for the U4/U6 di-snRNP to assemble onto the pre-mRNA, it must first associate with U5 snRNP to form the U4/U6•U5 tri-snRNP complex (Fig. 2.3). U5 snRNA can be divided into two major domains: the 5′ domain, which contains a complex stem loop structure, and the 3′ domain, which contains the single-stranded Sm protein binding site followed by a 3′ stem loop that varies in both size and sequence (Fig. 2.3; Guthrie and Patterson 1988). In *S. cerevisiae*, there are two functional forms of U5, a short and long form, with the short form terminating just prior to the 3′ stem loop structure (Patterson and Guthrie 1987). The 5′ stem loop is comprised of a long stem loop (loop 1) that is broken into three segments by two internal loops, IL1 and IL2, and a stem loop on the 5′ side of IL2 that is unique to *S. cerevisiae* (Guthrie and Patterson 1988). Loop 1, which makes direct contacts with the exon junction (Sontheimer and Steitz 1993; Newman et al. 1995), exhibits extreme sequence conservation where nine of eleven nucleotides are invariant across eukaryotes. *C. albicans* is an exception to this where two additional nucleotides have been reported to deviate from the loop 1 consensus sequence (Mitrovich and Guthrie 2007).

Free U5 snRNA associates with eight different proteins—in addition to the heptameric Sm protein ring—to form the free U5 snRNP; all of them have a mammalian homolog (Stevens et al. 2001; Fabrizio et al. 2009). While Brr2, Prp8, and the only known spliceosomal GTPase, Snu114 (Fabrizio et al. 1997), are the key players in spliceosome activation, they also appear to play a major role in U5 snRNP formation and stability (Dix et al. 1998). Prp8 physically contacts U5 snRNA on both sides of IL1 and IL2, as well as loop 1 of the 5′ stem loop, while Snu114 contacts the 5′ side of IL2 (Dix et al. 1998). The stability of Prp8 depends on its ability to interact with Snu114, and stable interaction between these proteins requires the binding, but not hydrolysis, of GTP by Snu114 (Brenner and Guthrie 2006). Stable association of the GTP-bound Snu114/Prp8 dimer with U5 snRNA is required for stable association of Brr2, which interacts directly with Prp8, but not

with U5 snRNA or Snu114 (Dix et al. 1998; Brenner and Guthrie 2006). Once fully formed, this free U5 snRNP particle then associates with the U4/U6 di-snRNP, and, upon addition of five other proteins, generates the U4/U6•U5 tri-snRNP complex (Fig. 2.3; Stevens et al. 2001).

Spliceosome Activation

Activation of the spliceosome requires large conformational and compositional changes that result in the loss of U1 and U4 snRNPs, the acquisition of the Nineteen Complex (NTC), and formation of the catalytic core of the spliceosome (Fig. 2.4; Cheng and Abelson 1987; Hoskins et al. 2011). Each of these occurrences is precisely regulated and has been described as an allosteric cascade, in which the execution of one event is dependent on a conformational change associated with the previous event (Brow 2002). In some cases, this is the exchange of a base-pairing partner for a mutually exclusive partner, and in other cases it involves a structural rearrangement in a protein that changes the accessibility of an interaction domain. The key drivers of these rearrangements during spliceosome activation are the U5 snRNP associated DExD-box RNA helicase proteins Prp28, which acts on the 5′ss/U1 snRNA interaction, and Brr2, which acts on the U4/U6 di-snRNA interaction (Fig. 2.4; Raghunathan and Guthrie 1998). The U5 snRNP proteins Prp8 and Snu114 regulate the activity of both helicases through a delicate and finely tuned feedback system (Brenner and Guthrie 2005; Small et al. 2006).

It is not yet clear how or what recruits the U4/U6•U5 tri-snRNP to the assembling spliceosome, although initial association of this complex probably involves protein–protein interactions between U1 and U5 snRNP proteins. Yeast two hybrid interactions have been reported between the U1 snRNP proteins Prp40 and Snp1 and the U5 snRNP proteins Prp8 and Brr2, respectively, suggesting that these interactions allow the tri-snRNP complex to dock with the pre-spliceosome (Abovich and Rosbash 1997; Fromont-Racine et al. 2000). Stable association of the tri-snRNP with the pre-spliceosome is guided largely by Prp8, which stabilizes an interaction between loop 1 of U5 snRNA and the exons (Dix et al. 1998). Prior to docking with the pre-spliceosome, Prp8 and Snu114 inhibit Prp28 and Brr2 helicase activity through a mechanism that is not well understood; however, contact between the tri-snRNP and U1 snRNP proteins upon docking induces a large structural rearrangement in the C-terminus of Prp8 that results in the activation of both helicases (Kuhn et al. 1999, 2002; Brenner and Guthrie 2005).

The first major rearrangement during spliceosome activation is the exchange of U1 snRNA for U6 snRNA at the 5′ss, a process that requires ATP hydrolysis and Prp28 (Fig. 2.4; Staley and Guthrie 1999). Under normal wild type conditions, the yeast protein yU1C stabilizes the U1 snRNA/5′ss duplex. Mutations that alter either yU1C or U1 snRNA in the 5′ss binding region, however, render Prp28 dispensable for splicing, and, indeed, for cell viability (Chen et al. 2001). Since

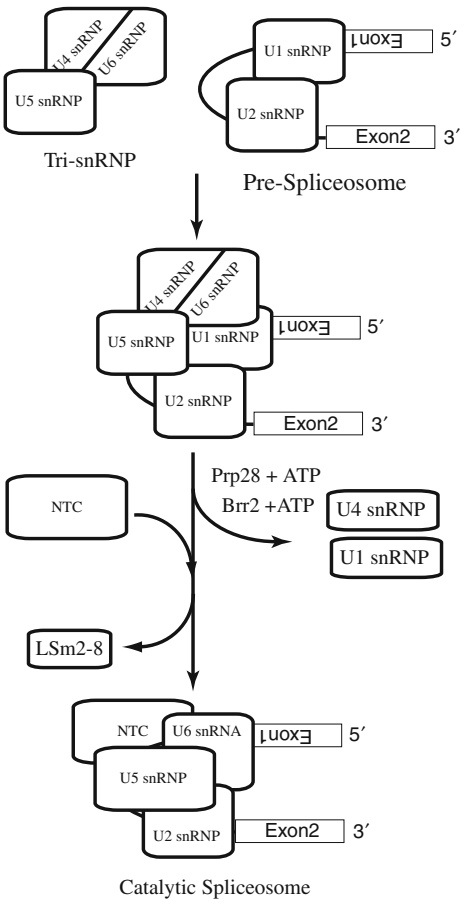


Fig. 2.4 Spliceosome activation, indicating tri-snRNP addition to the pre-spliceosome (*top*), U1 and U4 dissocation and NTC addition (*bottom*). SnRNP particles and protein complexes are indicated by *large rectangles*

these mutations act to destabilize the interaction between U1 snRNP and the 5′ss, Prp28 is thought to function as an antagonist to yU1C, destabilizing its interaction with the pre-mRNA to provide a more suitable environment for U6 snRNA binding to the 5′ ss (Chen et al. 2001). Formation of the U6 snRNA/5′ss duplex promotes the complete dissociation of U1 snRNP from the 5′ss (Kuhn et al. 1999; Chen et al. 2001). Notably, extending the U1 snRNA/5′ss interaction by several base pairs inhibits the switch for U6 snRNA and stalls spliceosome assembly (Staley and Guthrie 1999). This inhibition can be reversed by lengthening the U6 snRNA/5′ss interaction by several base pairs, suggesting that U1 and U6 compete for binding to the 5′ ss, resulting in an equilibrium between the two bound states (Staley and Guthrie 1999). Prp28 appears to play a role in proofreading the stability of the U6 snRNA/5′ss interaction, rejecting suboptimal 5′ss pre-mRNAs by sending them down a discard pathway (Yang et al. 2013).

A second major structural rearrangement during spliceosome activation is the disruption of the U4/U6 di-snRNP. Several lines of evidence suggest that

unwinding of the U4/U6 duplex is tightly coupled to destabilization of U1 snRNP at the 5′ss. First, when U1 snRNA/5′ss unwinding is blocked by extending base pairing, U4/U6 duplex unwinding is also blocked (Staley and Guthrie 1999). Second, in the presence of a mutation that extends stem I of the U4/U6 di-snRNA to include the 5′ss binding region of U6 snRNA, U4/U6 unwinding is impeded and U1 snRNP is retained in a stalled spliceosome assembly intermediate (Li and Brow 1996; Kuhn et al. 1999). A mutation in Prp8 is capable of suppressing the conditional phenotype generated by the stem I-lengthening mutation, suggesting that U4/U6 unwinding is triggered by Prp8 only after stable association of U6 snRNA with the 5′ss (Kuhn et al. 1999; Staley and Guthrie 1999). Such a system would ensure that catalytic structures do not form prior to correct identification of the 5′ss, ensuring splicing fidelity during first step catalysis (Staley and Guthrie 1999).

While Prp8 is involved in regulating U4/U6 unwinding, it is Brr2 that plays an active role in unwinding the duplex (Raghunathan and Guthrie 1998b). In the absence of ATP or in the presence of a mutation in the helicase domain of Brr2, U4/U6 unwinding is inhibited (Raghunathan and Guthrie 1998b; Maeder et al. 2009). Genetic studies have implicated Prp8 as a negative regulator of Brr2, and in recent years some of the details of the mechanism of regulation have begun to surface (Kuhn et al. 2002). Specifically, the RNase H-like domain of Prp8 interacts directly with U4 and U6 snRNAs in single-stranded regions adjacent to U4/U6 stem I, the same region of U4 that is required for loading Brr2 onto the duplex (Mozaffari-Jovin et al. 2012). Prp8 and Brr2 physically interact with the same region of U4 snRNA in a mutually exclusive manner, with Prp8 blocking U4/U6 unwinding by preventing Brr2 from binding (Mozaffari-Jovin et al. 2012). A high-resolution crystal structure has revealed that Prp8 further blocks Brr2 activity by inserting its C-terminal tail into the RNA binding tunnel of Brr2, inhibiting the ATP-dependent helicase activity of Brr2 (Mozaffari-Jovin et al. 2013).

Once Brr2 has loaded onto U4 snRNA, it translocates along U4 to unwind U4/U6 stem I (Hahn et al. 2012; Mozaffari-Jovin et al. 2013). It is not yet clear how stem II is unwound, since Brr2 would encounter the protein-bound U4 snRNA 5′ stem loop before reaching stem II. It is possible that Brr2 continues to translocate along U4 snRNA, displacing proteins as they are encountered, and finally unwinding stem II (Nielsen and Staley 2012). Alternatively, Brr2 might somehow jump the 5′ stem loop to immediately unwind stem II following stem I unwinding, or it might not be involved in stem II unwinding at all (Nielsen and Staley 2012). What is known is that following release of U4 snRNA, Brr2 activity must be turned off to allow formation of the catalytic center of the spliceosome. Snu114 appears to function as a regulator of Brr2, since Brr2 activity is repressed when Snu114 is bound to GDP (Small et al. 2006). Importantly, while GTP hydrolysis is not required for U4/U6 unwinding, it is required for U4 snRNA release from the assembling spliceosome (Bartels et al. 2003; Small et al. 2006). Thus, the hydrolysis of GTP following U4/U6 unwinding might trigger the release of the destabilized U1 and U4 snRNPs from the assembled spliceosome, likely by influencing the physical contacts between proteins at the core of the spliceosome.

Following the release of U1 and U4, the Nineteen Complex (NTC) is recruited to stabilize the assembled spliceosome during spliceosome activation (Fig. 2.4; Chan et al. 2003). The NTC is composed of Prp19 and at least seven other Prp19-associated proteins, which assemble into the pre-formed NTC prior to association with the spliceosome (Chen and Cheng 2012). Binding of the NTC results in the destabilization of the LSm complex of proteins from the 3′ tail of U6 snRNA, allowing these U6 snRNA nucleotides to interact with an intronic region of the substrate near the 5′ss (Chan and Cheng 2005). Crosslinks between U6 snRNA and the NTC component Cwc2 have led to the proposal that Cwc2 serves to link the NTC to the spliceosome (McGrail et al. 2009). Interestingly, Prp19 itself contains a ubiquitin ligase motif at its N-terminus, and might regulate aspects of the splicing cycle through its ability to add ubiquitin to various proteins (Ohi et al. 2003). Indeed, Prp19 has been shown to ubiquitinate the U4/U6-associated protein Prp3 in humans, influencing tri-snRNP stability (Song et al. 2010). Further, ubiquitin is necessary for splicing in yeast, as inhibition of ubiquitin's ability to interact with other proteins through ubiquitin mutation, or the presence of an inhibitory small molecule, reduces splicing by reducing tri-snRNP levels (Bellare et al. 2008).

The catalytic core of the spliceosome is formed through base pairing between the U2 and U6 snRNAs. Notably, these interactions are mutually exclusive with U4/U6 interactions, supporting the proposal that U4 snRNA acts as a negative regulator of U6 snRNA, masking U6 nucleotides so that catalytic features of the active site do not form prematurely (Brow and Guthrie 1989). Specifically, the stem I region of U6 base pairs to U2 snRNA to form U2/U6 helix I, and the stem II region of U6 folds back on itself to generate an intramolecular stem loop structure known as the 3′ ISL (Madhani and Guthrie 1992, Fortner et al. 1994). Interestingly, the C-terminal region of Prp8 influences U6 3′ ISL formation and/or stability, highlighting the importance of Prp8 throughout the splicing cycle (Kuhn et al. 2002). It is not yet clear whether U2/U6 helix I forms before, after, or at the same time as the U6 3′ ISL, however unwinding of U4/U6 stem I prior to stem II suggests that correct association of U2 and U6 snRNAs might be a prerequisite to stem II unwinding and 3′ ISL formation. Once these catalytically important structures have formed, the spliceosome is considered to be fully activated and ready to splice a substrate.

Catalytic Steps

The splicing reaction consists of two sequential transesterification reactions separated by a period of spliceosomal remodeling. In the first reaction, the 2′ hydroxyl of a bulged adenosine found in the branch site consensus sequence of the intron reacts with the phosphodiester bond at the 5′ss (Fig. 2.5a; Padgett et al. 1984; Konarska et al. 1985). This results in the formation of an unusual 2′–5′ phosphodiester linkage joining the 5′ end of the intron to the branch point adenosine,

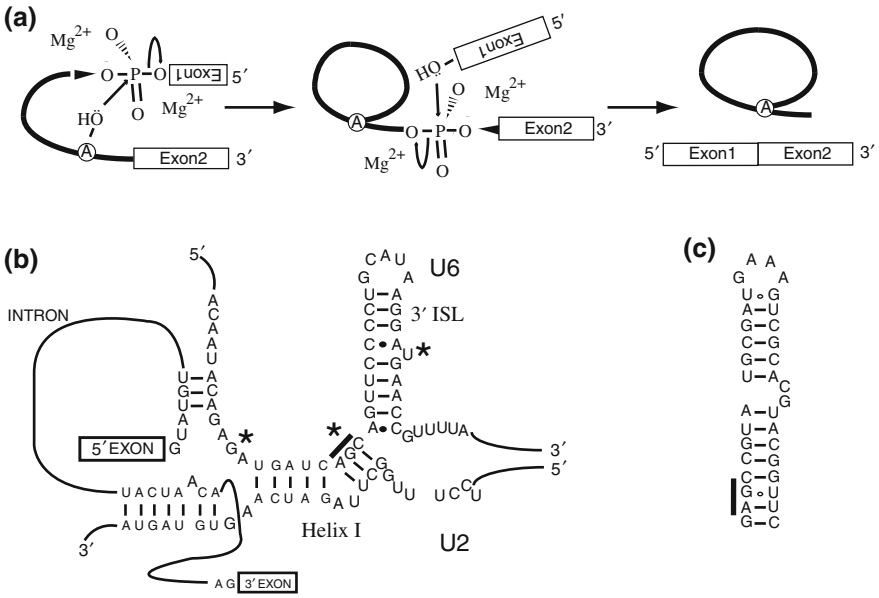


Fig. 2.5 **a** Nucleophillic reaction of the 2′-hydroxyl of the branchpoint adenosine (*circled*) with the 5′ splice site liberates the 5′ exon concomitantly with the lariat intron/3′ exon intermediate (*middle*). Nucleophillic reaction of the 3′-hydroxyl of the 5′ exon with the 3′ splice site results in the ligation of the exons and release of the intron in the form of a lariat (*right*). **b** RNA interactions between U6 (*above*), U2 (*below*), and pre-mRNA (*left*) at the first chemical step. Catalytic Mg^{2+} ion binding sites are indicated with an *asterisk*, and the AGC triad is marked with a *bar*. **c** Catalytic domain V of a group II intron. The AGC triad is marked with a *bar*

with concomitant liberation of the 5′ exon (Padgett et al. 1984; Konarska et al. 1985). In the second step, the 3′ hydroxyl group of the 5′ exon reacts with the phosphodiester bond at the 3′ss, joining the 5′ and 3′ exons through a standard 3′–5′ phosphodiester linkage, with concomitant release of the intron in the form of a lariat (Padgett et al. 1984; Konarska et al. 1985). Both chemical steps are inferred to proceed through an in-line S_N2 nucleophilic reaction based on the inversion of stereochemistry observed at the chiral phosphates (Maschhoff and Padgett 1993; Moore and Sharp 1993). While the spliceosome is composed largely of proteins, a long-held view is that the splicing reactions might actually be catalyzed by the highly conserved snRNAs located at the catalytic core of the spliceosome (Madhani and Guthrie 1992).

A catalytic function for U6 snRNA has been suspected for decades, not only because of its high level of sequence and size conservation, but also because of mechanistic and structural similarities to group II self-splicing introns (Madhani and Guthrie 1992; Peebles et al. 1995). At the active site of the spliceosome, U6 snRNA adopts a conformation that resembles the active domain of group II introns through formation of U2/U6 helix I and the U6 3′ ISL (Fig. 2.5b). U2/U6 helix I contains an invariant AGC triad that is required for exon ligation (Madhani and

Guthrie 1992; Fabrizio and Abelson 1992; Hilliker and Staley 2004; Lee et al. 2010). The AGC triad is also present in a base-paired structure in group II introns (Fig. 2.5c), and like the spliceosome, has a strict requirement for a purine at the second position (Peebles et al. 1995; Hilliker and Staley 2004). Interestingly, in both systems the AGC triad is less tolerant of mutation than the complementary nucleotides to which it is base paired, demonstrating an important role for these nucleotides in addition to base pairing (Madhani and Guthrie 1992; Peebles et al. 1995). Directly adjacent to U2/U6 helix I lies the 3′ISL, which, like domain V of group II introns, contains a small internal bulge on the 3′ side of the stem (Fabrizio and Abelson 1992; Lee et al. 2010; Peebles et al. 1995).

Mechanistically, pre-mRNA splicing and group II self-splicing are identical: both are Mg^{2+}-dependent processes that result in the removal of a lariat intron (Peebles et al. 1995; van der Veen et al. 1986; Cech 1986). Steitz and Steitz (1993) have proposed a two metal ion mechanism for these reactions in which one metal ion activates the sugar hydroxyl, while the other metal ion directly coordinates and stabilizes the oxyanion leaving group. To date, three Mg^{2+} ion binding sites have been identified in U6 snRNA: one in the AGC triad, one at position U80 in the internal loop of the 3′ ISL, and a third in the almost invariant ACAGAGA sequence, which base pairs to the 5′ splice site of the pre-mRNA transcript and is located 5′ of U2/U6 helix I (Fabrizio and Abelson 1992; Lee et al. 2010). In order for these Mg^{2+} ions to work in concert during the splicing reactions, the 3′ ISL must be closely juxtaposed with the 5′ splice site of the pre-mRNA transcript. Chemical structure probing of assembled spliceosomes has shown that this is indeed the case, with all three of these Mg^{2+} binding sites located in close proximity to position ten of the intron prior to the first catalytic step (Rhode et al. 2006). This constrains the structure of the active core in three dimensions, placing all three Mg^{2+} ions close to the reactive groups for the first chemical step. Following the first reaction, the accessibility of the 3′ ISL changes, supporting the view that some level of spliceosomal remodeling occurs between the two splicing reactions (Rhode et al. 2006).

While the exact role of the Mg^{2+} ion that is coordinated at each site has not yet been elucidated, Yean et al. (2000) showed that substitution of a phosphorothioate at position U80 in the 3′ ISL reconstitutes fully assembled, but catalytically inactive spliceosomes. Only in the presence of more thiophilic metal ions does splicing proceed, demonstrating that it is the splicing reaction, not spliceosome assembly, that requires this Mg^{2+} ion. While metal substitution fails to restore splicing in a phosphorothioate-substituted internal loop of a group II intron (Gordon and Piccirilli 2001), Tb^{3+} ion cleavage at this position suggests that this is indeed a site of metal ion coordination (Sigel et al. 2000). Notably, Fica et al. (2013) showed that U6 snRNA catalyzes both splicing reactions by positioning Mg^{2+} ions that are critical to stabilize the leaving groups, confirming Steitz and Steitz's (1993) original proposal. Further, a reaction that resembles pre-mRNA splicing has been performed in the presence of Mg^{2+} ions in a protein-free system consisting of regions of U2 and U6 snRNA that make up the proposed catalytic domain (Valadkhan and Manley 2001). Thus, the spliceosome

can be considered a metallo-enzyme in which U6 snRNA plays a key role in coordinating these metal ions.

In addition to metal ion coordination by snRNAs at the active site of the spliceosome, Mg^{2+} ions are coordinated by protein components, although a direct role in catalysis for these metal ions has not been shown. Prp8 is the largest spliceosomal protein (260 kDa), containing RNase H-like, endonuclease-like, and reverse transcriptase-like domains, none of which are catalytically active (Jackson et al. 1988; Pena et al. 2008; Dlakic and Mushegian 2011). Different first and second step conformations for Prp8 have been suspected for some time based on genetic findings, and recent structural work with human Prp8 has revealed a subtle difference in Prp8 conformation in which one state, an open form, allows Mg^{2+} ion coordination in the RNase H-like domain, while the other, the closed form, does not (Schellenberg et al. 2013). The Mg^{2+}-bound open state functions during the second catalytic step, where Mg^{2+} ion coordination was shown to promote exon ligation (Schellenberg et al. 2013). Schellenberg et al. (2013) suggest that Prp8 might present its Mg^{2+} ion at the active site along with two other metal ions presented by the snRNAs to generate a three-metal spliceosomal active site as observed for other enzymes that catalyze phosphoryl transfer reactions. In contrast, Abelson (2013) favors a role for this Mg^{2+} ion in stabilizing the second step active site conformation rather than a direct role in catalysis, given that the RNase H-like domain of Prp8 is catalytically inert.

Spliceosome Remodeling Between Catalytic Steps I and II

A general theme in spliceosome remodeling between the catalytic steps is beginning to emerge in which the components that are required for each step are present throughout both splicing reactions, but with substantial 'toggling' of these components to generate the appropriate active site for each step. For example, U2 snRNA toggles between two mutually exclusive stem structures: stem IIa and stem IIc (Fig. 2.6a). Stem IIc is required for catalysis of both steps of the splicing reaction, while stem IIa is required for spliceosome assembly and substrate rearrangement between the two catalytic steps (Hilliker et al. 2007). Thus, U2 toggles between these two conformations to allow spliceosome assembly and catalysis to proceed, and there is evidence to suggest that the RNA-dependent helicase Prp16 plays a role in this interchange (Fig. 2.6b; Hilliker et al. 2007). Similar events have been reported for active site protein components in which the affinity for protein binding in the spliceosome toggles between low and high affinity states. For example, Prp16 and Slu7 bind the activated spliceosome through low affinity entry sites that are converted to high affinity binding sites following the first catalytic step, when the action of these proteins is required (Ohrt et al. 2013).

As in spliceosome assembly and activation, several RNA-dependent ATPases are required to promote each splicing reaction, probably by facilitating the formation of the step one and step two active sites. Interaction of Prp2 with the intron

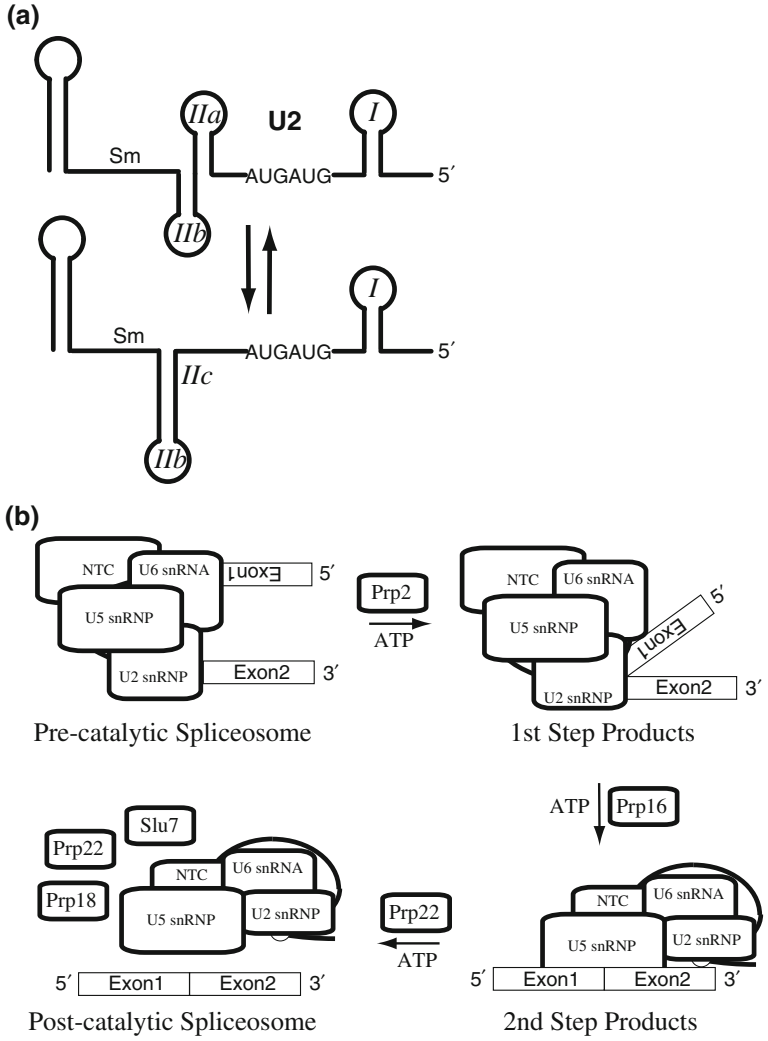


Fig. 2.6 **a** Conformational toggling in U2 snRNA, showing the switch between stem-loop IIa (*top*) and the mutually exclusive stem IIc (*bottom*), which lengthens stem-loop IIb. **b** ATP hydrolases of the helicase family associated with the chemical steps of splicing and spliceosome disassembly

prior to the first catalytic step is required for the splicing reactions to proceed (Fig. 2.6b), and, in addition to making direct contacts with the pre-mRNA, Prp2 interacts with the carboxyl-terminus of Brr2 (Liu and Cheng 2012). This interaction has been proposed to allow recruitment of Prp2 to the pre-catalytic spliceosome (Liu and Cheng 2012). Contact between Prp2 and Brr2 promotes the ATPase activity of Prp2, which results in the displacement of nine of eleven U2

snRNP-associated proteins (the SF3a and SF3b complexes in humans) through a mechanism that is not yet understood (Warkocki et al. 2009; Lardelli et al. 2010; Liu and Cheng 2012). The presence of the U2 snRNP-associated proteins at the BP region of the pre-mRNA may mask the reactive 2′-hydroxyl of the branchpoint adenosine until the spliceosome has correctly formed the step one active site. Removal of these proteins by Prp2 exposes the 2′-hydroxyl in a conformation that is compatible with in-line reaction with the phosphodiester bond at the 5′ss (Lardelli et al. 2010). Notably, these U2 snRNP proteins can be isolated in a particle containing U2 snRNA when purified spliceosomes are disassembled, suggesting that the U2 snRNP proteins, while displaced from the branchpoint for the first catalytic step, might remain loosely associated with the spliceosome throughout the splicing reactions (Fourmann et al. 2013).

Following the first catalytic step of splicing, the spliceosome re-positions the substrate for the second catalytic step, and the key driver of this remodeling event is the RNA-dependent ATPase, Prp16 (Fig. 2.6b; Schwer and Guthrie 1992). Prp16 is required specifically for the second catalytic step where it influences 3′ss cleavage and exon ligation, however, it has been shown to associate with the spliceosome in an ATP-independent manner prior to the first catalytic step to stabilize binding of the protein Cwc25 at the branchpoint (Schwer and Guthrie 1991; Tseng et al. 2011). Following the first catalytic step, Prp16 functions in an ATP-dependent manner to displace Yju2 and Cwc25 to allow for the association of the second step splicing factors Slu7, Prp18, and Prp22 (Tseng et al. 2011). Notably, Cwc25 is not displaced by Prp16 alone, but requires the stable association of Slu7 and Prp18, which are required to dock the 3′ss into the step two active site of the spliceosome (Ohrt et al. 2013). Interestingly, exon ligation can occur in the absence of Slu7 and Prp18 when the distance between the branchpoint and 3′ss is short, however, both proteins are required when this distance is longer than seven nucleotides (Brys and Schwer 1996; Ohrt et al. 2013).

In a genetic study, Mefford and Staley (2009) showed that Prp16 acts to destabilize U2/U6 helix I after the first catalytic step. However, since helix I integrity is important for both catalytic steps, they proposed that helix I reforms prior to second step catalysis. This is reminiscent of a second region of U2 snRNA discussed previously that undergoes toggling between the stem IIa and stem IIc conformations throughout the splicing cycle (Hilliker et al. 2007). Following 5′ss cleavage, Prp16 has been proposed to disrupt the stem IIc catalytic conformation by destabilizing stem IIc itself, as well as to destabilize interactions that are mutually exclusive with stem IIa, thereby promoting stem IIa formation (Hilliker et al. 2007). While the specific Prp16 substrate has yet to be identified, it is tempting to speculate that Prp16's role in displacing Yju2 and Cwc25 is an indirect consequence of Prp16 unwinding various U2 snRNA duplexes. Unwinding these structures would relax the catalytic core of the spliceosome, allowing for substrate re-positioning, while reformation of the snRNA structures would result in stable formation of the step two active site.

Once the splicing reactions have been completed, the mature mRNA product must be released from the spliceosome, and Prp22 is the RNA-dependent ATPase

responsible for promoting this event (Fig. 2.6b; Company et al. 1991; Schwer and Gross 1998). Like many of the ATPases encountered so far, Prp22 performs both an ATP-independent and an ATP-dependent function in splicing. The ATP-independent function is not well characterized, but is only required when the distance between the BP and the 3′ss is greater than 20 nucleotides (Schwer and Gross 1998; Schwer 2008). This function is required prior to execution of the second step, when Prp22 has been proposed to act in concert with Slu7 and Prp18 to position the 3′ss and 3′-hydroxyl of the 5′-exon for catalysis (Schwer 2008). Site-specific crosslinks and RNase H protection of the mRNA downstream of the exon-exon junction in the presence of Prp22 suggest that a conformational rearrangement following the second catalytic step places Prp22 on the mRNA at this location (Schwer 2008). Prp22 then acts to unwind the mRNA/U5 snRNA duplex, releasing the mRNA from the spliceosome using the energy generated through ATP hydrolysis (Schwer and Gross 1998; Schwer 2008). Following mRNA release, Slu7, Prp18, and Prp22 dissociate from the spliceosome (James et al. 2002).

Spliceosome Disassembly

After a substrate has been spliced, the spliceosome undergoes disassembly, resulting in the separation of U2, U5, U6, the NTC, and the lariat intron, and thereby allowing the spliceosomal components to be recycled for subsequent rounds of splicing (Fig. 2.7; Tsai et al. 2005). The DExD/H box RNA helicase Prp43, which associates with Ntr1 and Ntr2 to form the NTR complex, is responsible for promoting spliceosome disassembly in an ATP-dependent manner following mRNA release (Arenas and Abelson 1997; Tsai et al. 2005). Prp43 helicase activity is greatly enhanced through its interaction with Ntr1, demonstrating that Ntr1 is an accessory factor that is required to regulate Prp43 activity (Tanaka et al. 2007). Prp43 is recruited to the spliceosome through an interaction between Ntr2 and the U5 snRNP-component Brr2 (Tsai et al. 2007). Since Brr2 is present early in spliceosome assembly and throughout both catalytic steps, it is notable that binding of Ntr2 is competitively inhibited by the presence of Prp16 and Slu7, ensuring that spliceosome disassembly is not prematurely triggered through early association of the NTR complex with Brr2 (Chen et al. 2013).

Whether or not Brr2 helicase activity is required during intron release and spliceosome disassembly is up for debate. Small et al. (2006) reported that in a GTP-bound state, Snu114 derepresses Brr2 activity after the second catalytic step, resulting in intron release and spliceosome disassembly in much the same way as observed for U4/U6 unwinding during spliceosome assembly. This model presents another example of toggling throughout the splicing cycle, whereby hydrolysis of GTP to GDP results in repression of Brr2 activity following U4/U6 unwinding; subsequent exchange of the GDP for a new GTP following the splicing reactions derepresses Brr2 to allow spliceosome disassembly. Indeed, the RNA-dependent

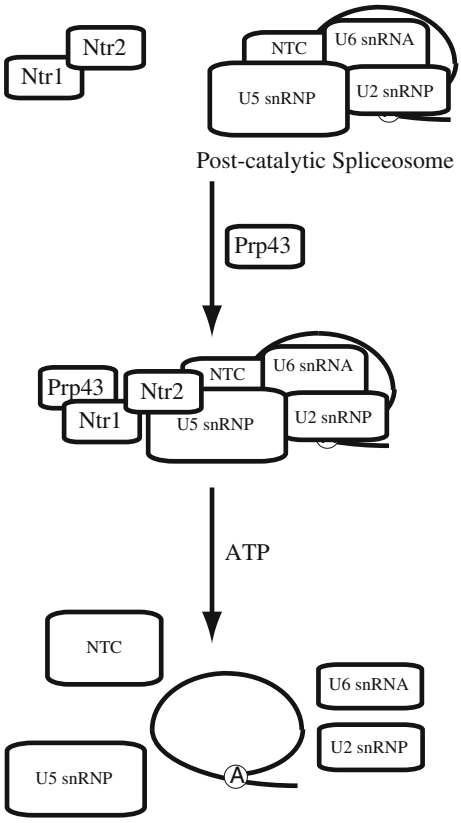


Fig. 2.7 Spliceosome disassembly mediated by Prp43 and the NTR complex. The excised intron is shown as a lariat with the BP adenosine *circled*. Proteins and snRNP particles are indicated by *rectangles*

ATPase activity of Brr2 is preferentially stimulated by annealed U2/U6, suggesting that the U2/U6 duplex could be a Brr2 substrate (Xu et al. 1996). However, Fourmann et al. (2013) recently showed that while Prp43 is necessary and sufficient for spliceosome disassembly, Brr2 is not required. Since Brr2 activity is specifically dependent on ATP hydrolysis, the fact that spliceosome disassembly proceeded as efficiently in the presence of UTP, CTP, and GTP as it did in the presence of ATP strengthens the argument that Brr2 activity is not required at this step (Fourmann et al. 2013).

The conflicting results reported by Small et al. (2006) and Fourmann et al. (2013) could reflect the different study systems used by the two groups. Fourmann et al. (2013) devised a purified splicing system with which stalled activated spliceosomes were isolated from an extract, followed by addition of recombinantly expressed and purified first and second step protein factors. The consequences of protein addition were then observed. In contrast, Small et al. (2006) used a tagged Prp43 to pull the Prp43-containing complex out of whole cell extract where potential endogenous factors reside that might play a role in splicing but have not yet been identified. It is possible that Prp43 activity destabilizes the spliceosome substantially, enough so that in the purified system, Brr2 activity is dispensable. In

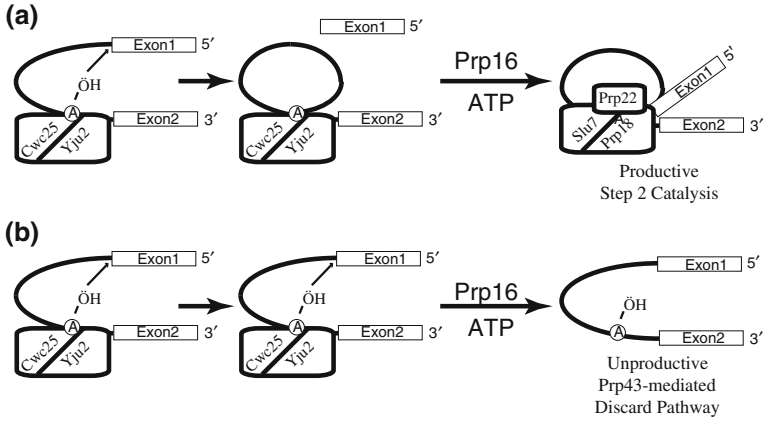


Fig. 2.8 Kinetic proofreading of the first catalytic step by Prp16. **a** An optimal substrate, in which reaction of the BP adenosine (*circled*) with the 5′ss is faster than ATP hydrolysis by Prp16, leading to dissociation of Cwc25 and Yju2 after the first chemical step has occurred. **b** A suboptimal pre-mRNA in which Prp16 hydrolysis occurs before the first chemical step, leading to premature dissociation of Cwc25 and Yju2, and subsequent Prp43-mediated disassembly

the absence of Brr2 activity, for example in the presence of UTP, the workload for Prp43 might increase to complete disassembly. In the Small et al. (2006) complex, other factors might contribute to the stability of the disassembling spliceosome such that Brr2 activity is required for efficient disassembly. Further experimentation will be required to reconcile these differences.

Splicing Fidelity

Since pre-mRNA splicing involves the removal of intervening sequences and ligation of protein coding sequences to generate a continuous translation template, splicing must proceed with single nucleotide precision to avoid introducing nucleotide insertions or deletions that would result in the translation of frame-shifted, aberrant products. The spliceosome has evolved a number of proofreading mechanisms to ensure fidelity throughout assembly and catalysis. In these, the spliceosome acts to promote splicing of optimal substrates, while antagonizing splicing of suboptimal substrates (Semlow and Staley 2012). One proofreading mechanism for which there is support in splicing is kinetic proofreading, originally described independently by Hopfield (1974) and Ninio (1975) in the translation field. In splicing, kinetic proofreading has been observed in early spliceosome assembly, where U2 snRNP association with the branchpoint, and exchange of U1 for U6 at the 5′ss, are proofread by Prp5 and Prp28, respectively. Other examples have been found through the first and second catalytic steps (Xu and Query 2007; Yang et al. 2013; Burgess and Guthrie 1993; Mayas et al. 2006).

In the kinetic proofreading model, energy is expended to allow for inspection of the substrate before allowing the substrate to proceed down a productive pathway. Optimal substrates undergo reaction quickly, while the time required for reaction of suboptimal substrates is longer (Fig. 2.8). Several splicing ATPases, such as Prp16 and Prp22, have been implicated as "timers" during these proofreading stages, in which splicing of optimal substrates proceeds more rapidly than the ATPase acts (Fig. 2.8a; Burgess and Guthrie 1993; Mayas et al. 2006). As a consequence, hydrolysis of ATP promotes a conformational change that shuffles the substrate down a productive pathway. However, when suboptimal substrates are encountered, ATP hydrolysis occurs more rapidly than the splicing reaction (Fig. 2.8b). This results in a conformational change in the spliceosome that promotes the rejection of the substrate through a discard pathway. Discrimination between fast and slow substrates may be based in part on the spliceosome's ability to discriminate between substrates that are positioned correctly for the chemical steps and those that are not (Chua and Reed 1999).

The role of Prp16 in kinetic proofreading during the first catalytic step has been well characterized and serves as an excellent example of proofreading by the spliceosome. Proofreading at this stage involves kinetic competition between the Prp16-dependent release of Cwc25 and the first transesterification reaction (Fig. 2.8a; Tseng et al. 2011). When the splicing machinery encounters an optimal substrate, the transesterification reaction proceeds more rapidly than the removal of Cwc25, and thus Cwc25 is displaced by Prp16 after the first catalytic step, thereby making way for second-step splicing factors. In the case of a suboptimal substrate containing branchpoint mutations, however, ATP hydrolysis by Prp16 occurs more rapidly than the transesterification reaction, resulting in the premature release of Yju2 and Cwc25 from the spliceosome prior to completion of the first transesterification reaction (Fig. 2.8b; Tseng et al. 2011). Discard of the suboptimal substrate at this point involves the disassembly factor Prp43 (Koodathingal et al. 2013). In fact, Prp43-mediated spliceosome disassembly can be initiated after the action of Prp2, Prp16, or Prp22, following their dissociation from the spliceosome when a suboptimal substrate is encountered. This suggests that Prp43 plays a more general role in discarding suboptimal substrates throughout catalysis, in addition to disassembling the spliceosome following splicing of an optimal substrate (Chen et al. 2013).

Concluding Remarks

Over the last several decades, much work has been completed to understand the chemical mechanism of the splicing reactions and the composition of the spliceosome, which is responsible for catalyzing these reactions. Despite this wealth of information, very little is known about the exact role of many splicing factors, and even less is known about the mechanisms through which these factors function. With recent advances in the technology used to study splicing, we now have an

opportunity to investigate and explore questions that could not be addressed previously. For example, we are seeing a shift from analyzing bulk splicing in whole cell extract to monitoring the fate of individual substrates by fluorescence microscopy. As a result of this transition, we are already beginning to understand the kinetics of individual steps, and the order of association and dissociation events, with greater depth. These types of inquiries, along with progress in atomic-resolution structure determination of splicing complexes, will lead to a better understanding of the intricate details of the splicing cycle.

Acknowledgments This work was supported by NSERC Discovery Grant 298521 to SDR and an NSERC PGS award to EAD, as well as by awards from UNBC's Office of Research.

References

Abelson J (2013) Toggling in the spliceosome. Nat Struct Mol Biol 20:645–647

Abovich N, Liao XC, Robash M (1994) The yeast MUD2 protein: an interaction with PRP11 defines a bridge between commitment complexes and U2 snRNP addition. Genes Dev 8:843–854

Abovich N, Rosbash M (1997) Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. Cell 89:403–412

Arenas JE, Abelson J (1997) Prp43: an RNA helicase-like factor involved in spliceosome disassembly. PNAS 94:11798–11802

Ares M Jr, Grate L, Pauling MH (1999) A handful of intron-containing genes produces the lion's share of yeast mRNA. RNA 5:113–1139

Bartels C, Urlaub H, Lührmann R et al (2003) Mutagenesis suggests several roles of Snu114p in pre-mRNA splicing. J Biol Chem 278:28324–28334

Bell M, Schreiner S, Damainov A et al (2002) p110, a novel human U6 snRNP protein and U4/U6 snRNP recycling factor. EMBO J 21:2724–2735

Bellare P, Small EC, Huang X et al (2008) A role for ubiquitin in the spliceosome assembly pathway. Nat Struct Mol Biol 15:444–451

Bon E, Casaregola S, Blandin G et al (2003) Molecular evolution of eukaryotic genomes: hemiascomycetous yeast spliceosomal introns. NAR 31:1121–1135

Brenner TJ, Guthrie C (2005) Genetic analysis reveals a role for the C terminus of the *Sacchaomyces cerevisiae* GTPase Snu114 during spliceosome activation. Genetics 170:1063–1080

Brenner TJ, Guthrie C (2006) Assembly of Snu114 into U5 snRNP requires Prp8 and a functional GTPase domain. RNA 12:862–871

Brody E, Abelson J (1985) The "Spliceosome": Yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. Science 228:963–967

Brow DA, Guthrie C (1988) Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. Nature 334:213–218

Brow DA, Guthrie C (1989) Splicing a spliceosomal RNA. Nature 337:14–15

Brow DA (2002) Allosteric cascade of spliceosomal activation. Annu Rev Genet 36:333–360

Brys A, Schwer B (1996) Requirement for Slu7 in yeast pre-mRNA splicing is dictated by the distance between the branchpoint and 3′ splice site. RNA 2:707–717

BurgessSM Guthrie C (1993) A mechanism to enhance mRNA splicing fidelity: the RNA-dependent ATPase Prp16 governs usage of a discard pathway for aberrant lariat intermediates. Cell 73:1377–1391

Burset M, Seledtov IA, Solovyev VV (2000) Analysis of canonical and non-canonicle splice sites in mammalian genomes. NAR 28:4364–4375
Cech T (1986) The generality of self-splicing RNA: relationship to nuclear mRNA splicing. Cell 44:207–210
Chan SP, Kao DI, Tsai WY et al (2003) The Prp19p-associated complex in splicesome activation. Science 302:279–282
Chan SP, Cheng SC (2005) The Prp19-associated complex is required for specifying interactions of U5 and U6 with the pre-mRNA during spliceosome activation. J Biol Chem 280:31190–31199
Chen HC, Cheng SC (2012) Functional roles of protein splicing factors. Biosci Rep 32:345–359
Chen HC, Tseng CK, Tsai RT et al (2013) Link of NTR-mediated spliceosome disassembly with DEAH-box ATPases Prp2, Prp16, and Prp22. Mol Cell Biol 33:514–525
Chen JY, Stands L, Staley JP et al (2001) Specific alterations of U1-C protein or U1 small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD box splicing factor. Mol Cell 7:227–232
Cheng SC, Abelson J (1987) Spliceosome assembly in yeast. Genes Dev 1:1014–1027
Chua K, Reed R (1999) The RNA splicing factor hSlu7 is required for correct 3′ splice site choice. Nature 402:207–210
Collemare J, van der Burget A, de Wit PJ (2013) At the origin of spliceosomal introns: is multiplication of intron-like elements the main mechanism of intron gain in fungi? Commun. Integr. Biol. 6(2):e23147
Company M, Arenas J, Abelson J (1991) Requirement of the RNA helicase-like Prp22 for release of messenger RNA from spliceosomes. Nature 349:487–493
Coulombe-Huntington J, Majewski J (2007) Characterization of intron loss events in mammals. Genome Res 17:23–32
Crawford DJ, Hoskins AA, Friedman LJ et al (2013) Single-molecule colocalization FRET evidence that spliceosome activation precedes stable approach of 5′ splice site and branch site. PNAS 110:6783–6788
Deutsch M, Long M (1999) Intron-exon structures of eukaryotic model organisms. NAR 27:3219–3228
Dlakic M, Mushegian A (2011) Prp8, the pivotal protein of the spliceosomal catalytic center, evolved from a retroelement-encoded reverse transcriptase. RNA 17:799–808
Dix I, Russell CS, O'Keefe RT et al (1998) Protein–RNA interactions in the U5 snRNP of *Saccharomyces ceravisiae*. RNA 4:1675–1686
Fabrizio P, Abelson J (1992) Thiophosphates in yeast U6 snRNA specifically affect pre-mRNA splicing in vitro. NAR. 20:3659–3664
Fabrizio P, Laggerbauer B, Lauber J et al (1997) An evolutionarily conserved U5 snRNP-specific protein is a GTP-binding factor closely related to the ribosomal translocase EF-2. EMBO 16:4092–4106
Fabrizio P, Dannenberg J, Dube P et al (2009) The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. Mol Cell 36:593–608
Farlow A, Meduri E, Schlötterer C (2011) DNA double-strand break repair and the evolution of intron density. Trends Genet 27:1–6
Fica SM, Tuttle N, Novak T et al (2013) RNA catalyses nuclear pre-mRNA splicing. Nature 503:229–234
Fink GR (1987) Pseudogenes in yeast? Cell 49:5–6
Fortner DM, Troy RG, Brow DA (1994) A stem/loop in U6 RNA defines a conformational switch required for pre-mRNA splicing. Genes Dev 8:221–233
Fourmann JB, Schmitzová J, Christian H et al (2013) Dissection of the factor requirements for spliceosome disassembly and the elucidation of its dissociation products using a purified splicing system. Genes Dev 27:413–428
Fromont-Racine M, Mayes AE, Brunet-Simon A et al (2000) Genome-wide protein interaction screens reveal functional networks involving Sm-like proteins. Yeast 17:95–110

Gordon PM, Piccirilli JA (2001) Metal ion coordination by the AGC triad in domain 5 contributes to group II intron catalysis. Nat Struct Biol 8:893–898

Gozani O, Feld R, Reed R (1996) Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. Genes Dev 10:233–243

Gozani O, Potashkin J, Reed R (1998) A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. Mol Cell Biol 18:4752–4760

Grabowski PJ, Seiler SR, Sharp PA (1985) A multicomponent complex is involved in the splicing of messenger RNA precursors. Cell 42:345–353

Guthrie C, Patterson B (1988) Spliceosomal snRNAs. Annu Rev Genet 22:387–419

Hahn D, Kudla G, Tollervey D et al (2012) Brr2p-mediated conformational rearrangements in the spliceosome during activation and substrate repositioning. Genes Dev 26:2408–2421

Hilliker AK, Staley JP (2004) Multiple functions for the invariant AGC triad of U6 snRNA. RNA 10:921–928

Hilliker AK, Mefford MA, Staley JP (2007) U2 toggles iteratively between the stem IIa and stem IIc conformations to promote pre-mRNA splicing. Genes Dev 21:821–834

Hopfield JJ (1974) Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. PNAS 71:4135–4139

Hoskins AA, Friedman LJ, Gallagher SS et al (2011) Ordered and dynamic assembly of single spliceosomes. Science 331:1289–1295

Jackson SP, Lossky M, Beggs JD (1988) Cloning of the RNA8 gene of *Saccharomyces cerevisiae*, detection of the RNA8 protein, and demonstration that it is essential for nuclear pre-mRNA splicing. Mol Cell Biol 8:1067–1075

James SA, Turner W, Schwer B (2002) How Slu7 and Prp18 cooperate in the second step of yeast pre-mRNA splicing. RNA 8:1068–1077

Jurica MS, Moore MJ (2003) Pre-mRNA splicing: awash in a sea of proteins. Mol Cell 12:5–14

Käufer NF, Potashkin J (2000) Analysis of the splicing machinery in fission yeast: a comparison with budding yeast and mammals. NAR 28:3003–3010

Kistler AL, Guthrie C (2001) Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for Sub2, an essential spliceosomal ATPase. Genes Dev 15:42–49

Konarska MM, Grabowski PJ, Padgett RA et al (1985) Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors. Nature 313:552–557

Koodathingal P, Novak T, Piccirilli JA et al (2013) The DEAH-box ATPase Prp16 and Prp43 cooperate to proofread 5′ splice site cleavage during pre-mRNA splicing. Mol Cell 39:385–395

Kuhn AN, Li Z, Brow DA (1999) Splicing factor Prp8 governs U4/U6 RNA unwinding during activation of the spliceosome. Mol Cell 3:65–75

Kuhn AN, Reichl EM, Brow DA (2002) Distinct domains of splicing factor Prp8 mediate different aspects of spliceosome activation. PNAS 99:9145–9149

Kwan SS, Brow DA (2005) The N- and C-terminal RNA recognition motifs of splicing factor Prp24 have distinct functions in U6 binding. RNA 11:808–820

Lardelli RM, Thompson JX, Yates JR et al (2010) Release of SF3 from the intron branchpoint activates the first catalytic step of pre-mRNA splicing. RNA 16:516–528

Lee C, Jaladat Y, Mohammadi A et al (2010) Metal binding and substrate positioning by evolutionarily invariant U6 sequences in catalytically active protein-free snRNAs. RNA 16:2226–2238

Legrain P, Séraphin B, Robash M (1985) Early commitment of yeast pre-mRNA to the spliceosome pathway. Mol Cell Biol 8:3755–3760

Li Z, Brow D (1996) A spontaneous duplication in U6 spliceosomal RNA uncouples the early and late functions of the ACAGA element in vivo. RNA 2:879–894

Liu HL, Cheng SC (2012) The interaction of Prp2 with a defined region of the intron is required for the first splicing reaction. Mol Cell Biol 32:5056–5066

Madhani HD, Guthrie C (1992) A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. Cell 71:803–817

Maeder C, Kutach AK, Guthrie C (2009) ATP-dependent unwinding of U4/U6 snRNAs by the Brr2 helicase requires the C terminus of Prp8. Nat Struct Mol Biol 16:42–48
Maschhloff K, Padgett R (1993) The stereochemical course of the first step of pre-mRNA splicing. NAR. 21:5456–5462
Mayas RM, Maita H, Staley JP (2006) Exon ligation is proofread by the DExD/H-box ATPase Prp22p. Nat Struct Mol Biol 13:482–490
Mayes AE, Verdone L, Legrain P et al (1999) Characterization of Sm-like proteins in yeast and their association with U6 snRNA. EMBO J 18:4321–4331
McGrail JC, Krause A, O'Keefe RT (2009) The RNA binding protein Cwc2 interacts directly with the U6 snRNA to link the nineteen complex to the spliceosome during pre-mRNA splicing. NAR 37:4205–4217
Mefford MA, Staley JP (2009) Evidence that U2/U6 helix I promotes both catalytic steps of pre-mRNA splicing and rearranges in between these steps. RNA 15:1386–1397
Mekouar M, Blanc-Lenfle I, Ozanne C et al (2010) Detection and analysis of alternative splicing in Yarrowia lipolytica reveal structural constraints facilitating nonsense-mediated decay of intron-retaining transcripts. Genome Biol 11:R65. doi:10.1186/gb-2010-11-6-r65
Mitrovich QM, Guthrie C (2007) Evolution of small nuclear RNAs in *S. cerevisiae*, *C. albicans*, and other hemiascomycetous yeasts. RNA 13:2066–2080
Moore MJ, Sharp PA (1993) Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. Nature 365:364–368
Mozaffari-Jovin S, Santos KF, Hsiao HH et al (2012) The Prp8 RNase H-like domain inhibits Brr2-mediated U4/U6 snRNA unwinding by blocking Brr2 loading onto the U4 snRNA. Genes Dev 26:2422–2434
Mozaffari-Jovin S, Wandersleben T, Santos KF et al (2013) Inhibition of RNA helicase Brr2 by the C-terminal tail of the spliceosomal protein Prp8. Science 341:80–84
Neuvéglise C, Marck C, Gaillardin C (2011) The intronome of budding yeasts. C R Biol 34:662–670
Newman AJ, Teigelkamp S, Beggs JD (1995) snRNA interactions at 5′ and 3′ splice sites monitored by photoactivated crosslinking in yeast spliceosomes. RNA 1:968–980
Nielsen KH, Staley JP (2012) Spliceosome activation: U4 is the path, stem I is the goal, and Prp8 is the keeper. Let's cheer for the ATPase Brr2! Genes Dev 26:2461–2467
Ninio J (1975) Kinetic amplification of enzyme discrimination. Biochimie 57:587–595
O'Day CL, Dalbadie-McFarland G, Abelson J (1996) The *Saccharomyces cerevisiae* Prp5 protein has RNA-dependent ATPase activity with specificity for U2 small nuclear RNA. J Biol Chem 271:33261–33267
Ohi MD, Vander Kooi CW, Rosenberg JA et al (2003) Structural insights into the U-box, a domain associated with multi-ubiquitination. Nat Struct Biol 10:250–255
Ohrt T, Odenwälder P, Dannenberg J et al (2013) Molecular dissection of step 2 catalysis of yeast pre-mRNA splicing investigated in a purified system. RNA 19:902–915
Oubridge C, Ito N, Evans PR et al (1994) Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. Nature 372:432–438
Padgett RA, Konarska MM et al (1984) Lariat RNAs as intermediates and products in the splicing of messenger RNA precursors. Science 225:898–903
Parker R, Siliciano PG, Guthrie C (1987) Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. Cell 49:229–239
Patterson B, Guthrie C (1987) An essential yeast snRNA with a U5-like domain is required for splicing in vivo. Cell 49:613–624
Peebles CL, Zhang M, Perlman PS et al (1995) Catalytically critical nucleotides in domain 5 of a group II intron. PNAS 92:4422–4426
Pena V, Rozov A, Fabrizio P et al (2008) Structure and function of an RNAse H domain at the heart of the splicesome. EMBO J 27:2929–2940
Puig O, Gottschalk A, Fabrizio P et al (1999) Interaction of the U1 snRNP with nonconserved intronic sequences affects 5′ splice site selection. Genes Dev 13:569–580

Puig O, Bragado-Nilsson E, Koski T et al (2007) The U1 snRNP-associated factor Luc7p affects 5′ splice site selection in yeast and human. NAR 35:5874–5885
Rader SD, Guthrie C (2002) A conserved Lsm-interaction motif in Prp24 required for efficient U4/U6 di-snRNP formation. RNA 8:1378–1392
Raghunathan PL, Guthrie C (1998a) A spliceosomal recycling factor that reanneals U4 and U6 small nuclear ribonucleoprotein particles. Science 29:857–860
Raghunathan PL, Guthrie C (1998b) RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. Curr Biol 8:847–855
Rhode BM, Hartmuth K, Westhof E et al (2006) Proximity of conserved U6 and U2 snRNA elements to the 5′ splice site region in activated spliceosomes. EMBO J 25:2475–2486
Rymond BC, Rosbash M (1985) Cleavage of 5′ splice site and lariat formation are independent of 3′ splice site in yeast mRNA splicing. Nature 317:735–737
Sakharkar K, Chow V, Kangueane P (2004) Distribution of exons and introns in the human genome. In Silico Biol 4:387–393
Schellenberg MJ, Wu T, Ritchie DB et al (2013) A conformational switch in Prp8 mediates metal ion coordination that promotes pre-mRNA exon ligation. Nat Struct Mol Biol 20:728–724
Schwer B, Guthrie C (1991) Prp16 is an RNA-dependent ATPase that interacts transiently with the spliceisome. Nature 349:494–499
Schwer B, Guthrie C (1992) A conformational rearrangement in the spliceosome is dependent on Prp16 and ATP hydrolysis. J EMBO 11:5033–5039
Schwer B, Gross CH (1998) Prp22, a DExD-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. EMBO J 17:2086–2094
Schwer B (2008) A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. Mol Cell 30:743–754
Schwer B, Chang J, Shuman S (2013) Structure-function analysis of the 5′ end of yeast U1 snRNA highlights genetic interactions with the Msl5•Mud2 branchpoint-binding complex and other spliceosome assembly factors. NAR 41:7485–7500
Semlow DR, Staley JP (2012) Staying on message: Ensuring fidelity in pre-mRNA splicing. Trends Biochem Sci 37:263–273
Shannon KW, Guthrie C (1991) Suppressors of a U4 snRNA mutation define a novel U6 snRNP protein with RNA-binding motifs. Genes Dev 5:773–785
Shuster EO, Guthrie C (1988) Two conserved domains of yeast U2 snRNA are separated by 945 nonessential nucleotides. Cell 55:41–48
Shuster EO, Guthrie C (1990) Human U2 snRNA can function in pre-mRNA splicing in yeast. Nature 345:270–273
Sigel RK, Vaidya A, Pyle AM (2000) Metal ion binding sites in a group II intron core. Nat Struct Biol 7:1111–1116
Siliciano PG, Guthrie C (1988) 5′ splice site selection in yeast: genetic alterations in base-pairing with U1 reveal additional requirements. Genes Dev 2:1258–1267
Small EC, Leggett SR, Winans AA et al (2006) The EF-G-like GTPase Snu114p regulates spliceosome dynamics mediated by Brr2p, a DExD/H box ATPase. Mol Cell 23:389–399
Song EJ, Werner SL, Neubauer J et al (2010) The Prp19 complex and the $Usp4^{Sart3}$ deubiquitinating enzyme control reversible ubiquitination at the spliceosome. Genes Dev 24:1434–1447
Sontheimer EJ, Steitz JA (1993) The U5 and U6 small nuclear RNAs as active site components of the spliceosome. Science 262:1989–1996
Soochin C, Suk-Won J, Cohen A et al (2004) A phylogeny of *Caenorhabditis* reveals frequent loss of introns during nematode evolution. Genome Res 14:1207–1220
Staley JP, Guthrie C (1999) An RNA switch at the 5′ splice site requires ATP and the DEAD box protein Prp28p. Mol Cell 3:55–64
Steitz TA, Steitz JA (1993) A general two-metal-ion mechanism for catalytic RNA. PNAS 90:6498–6502

Stevens SW, Barta I, Ge HY et al (2001) Biochemical and genetic analyses of the U5, U6, and U4/U6•U5 small nucler ribonucleoproteins from *Saccharomyces cerevisiea*. RNA 7:1543–1553

Stevens SW, Ryan DE, Ge HY et al (2002) Composition and functional characterization of the yeast spliceosomal penta-snRNP. Mol Cell 9:31–44

Spingola M, Grate L, Haussler D et al (1999) Genome-wide bioinformatics and molecular analysis of introns in *Saccharomyces cerevisiae*. RNA 5:221–234

Surowy CS, van Santen VL, Scheib-Wixted SM et al (1989) Direct, sequence specific binding of the human U1-70 k ribonucleoprotein antigen protein to loop I of U1 small nuclear RNA. Mol Cell Biol 9:4179–4186

Tanaka N, Aronova A, Schwer B (2007) Ntr1 activates the Prp43 helicase to trigger release of lariat-intron from the spliceosome. Genes Dev 21:2312–2325

Tsai RT, Fu RH, Yeh FL et al (2005) Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2. Genes Dev 19:2991–3003

Tsai RT, Tseng CK, Lee PJ et al (2007) Dynamic interactions of Ntr1-Ntr2 with Prp43 and with U5 govern the recruitment of Prp43 to mediate spliceosome disassembly. J Mol Cell Biol 27:8027–8037

Tseng CK, Liu HL, Cheng SC (2011) DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. RNA 17:145–154

Valadkhan S, Manley J (2001) Splicing-related catalysis by protein-free snRNAs. Nature 413:701–707

Van der Veen R, Arnberg A, van der Horst G et al (1986) Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing in vitro. Cell 44:225–234

Venter JC, Adams MD, Myers EW et al (2001) The sequence of the human genome. Science 291:1304–1351

Vidovic I, Nottrott S, Hartmuth K et al (2000) Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. Mol Cell 6:1331–1342

Vinogradov AE (2001) Intron length and codon usage. J Mol Evol 52:2–5

Wang Q, Zhang L, Lynn B et al (2008) A BBP-Mud2p heterodimer mediates branchpoint recognition and influences splicing substrate abundance in budding yeast. NAR 36:2787–2798

Warkocki Z, Odenwälder P, Schmitzová J et al (2009) Reconstitution of both steps of *Saccharomyces cerevisiae* splicing with purified spliceosomal components. Nat Struct Mol Biol 16:1237–1243

Weber S, Aebi M (1988) In vitro splicing of mRNA precursors: 5′ cleavage site can be predicted from the interaction between the 5′ splice site region and the 5′ terminus of U1 snRNA. NAR 16:471–486

Wiest DK, O'Day CL, Abelson J (1996) In vitro studies of the Prp9•Prp11•Prp21 complex indicate a pathway for U2 small nuclear ribonucleoprotein activation. J Biol Chem 271:33268–33276

Wood V, Gwilliam R, Rajandream MA et al (2002) The genome sequence of *Schizosaccharomyces pombe*. Nature 415:871–880

Wu S, Romfo CM, Nilsen TW et al (1999) Functional recognition of the 3′ splice site AG by the splicing factor U2AF35. Nature 402:832–835

Xu D, Nouraini S, Field D et al (1996) An RNA-dependent ATPase associated with U2/U6 snRNAs in pre-mRNA splicing. Nature 381:709–713

Xu YZ, Query CC (2007) Competition between the ATPase Prp5 and branch region U2 snRNA pairing modulates the fidelity of spliceosome assembly. Mol Cell 28:838–849

Yang F, Wang X, Zhang ZM et al (2013) Splicing proofreading at 5′ splice sites by ATPase Prp28p. NAR 41:4660–4670

Yean SL, Wuenschell G, Termini J et al (2000) Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. Nature 408:881–884

Zhu T, Niu DK (2013) Mechanisms of intron loss and gain in the fission yeast *Schizosaccharomyces*. PLoS ONE 8:e61683

Chapter 3
Fungal Pre-mRNA 3′-End Processing

Aurelia Vavasseur and Yongsheng Shi

Abstract 3′ end processing of messenger RNAs (mRNAs) is not only an essential step in eukaryotic gene expression, but it also impacts many other aspects of mRNA maturation and decay. A large portion of eukaryotic genes produce multiple mRNAs with different 3′ ends through alternative cleavage/polyadenylation (APA). mRNA 3′ processing and especially APA has been increasingly recognized as an important mechanism for gene regulation. Much of what we currently know about eukaryotic mRNA 3′ processing came from studies using the genetically tractable yeast systems. Here we review the fungal mRNA 3′ processing system by describing both the evolutionarily conserved mechanisms as well as the fungus-specific features.

Contents

A. Vavasseur · Y. Shi (✉)
Department of Microbiology and Molecular Genetics, School of Medicine,
University of California, Irvine, CA 92697, USA
e-mail: yongshes@uci.edu

A. Sesma and T. von der Haar (eds.), *Fungal RNA Biology*,
DOI: 10.1007/978-3-319-05687-6_3,

Pre-mRNA 3′-End Processing: An Overview

The vast majority of eukaryotic mRNAs have a polyadenosine (poly(A)) tail at their 3′ ends (Chan et al. 2011; Zhao et al. 1999; Colgan and Manley 1997; Proudfoot 2011). The poly(A) sequences are not encoded in the genome, but are added posttranscriptionally through two chemical reactions, an endonucleolytic cleavage and the addition of a string of adenosines by the poly(A) polymerase (PAP). The poly(A) tails are critical for mRNA export, stability, and translation (Zhao et al. 1999; Colgan and Manley 1997; Chan et al. 2011; Proudfoot 2011). The 3′ end formation process itself is required for transcription termination and it significantly impacts other mRNA processing steps, including splicing (Kim et al. 2004; West et al. 2004; Connelly and Manley 1988; see Chap. 2 for more details). Mutations that disrupt the mRNA 3′ processing of critical genes and mutations in mRNA 3′ processing factors cause a number of diseases, including thalassemias, thrombophilia, and oculopharyngeal muscular dystrophy (Danckwardt et al. 2008; Chan et al. 2011).

Pre-mRNA 3′ end formation involves the assembly of several multisubunit complexes on specific cis-element that defines the polyadenylation site (PAS) (Chan et al. 2011; Shi et al. 2009; Skolnik-David et al. 1987; Humphrey et al. 1987). The majority of the 3′ processing factors are conserved throughout eukaryotic evolution (Darmon and Lutz 2012; Mandel et al. 2008). Interestingly, however, there are also a number of lineage-specific essential 3′ processing factors (Zhao et al. 1999). The loss and/or gain of these factors during evolution might be correlated with changes in the cis-element of PAS as described below.

APA is the phenomenon in which a gene can produce multiple mRNA isoforms with distinct 3′ ends through using alternative PAS (Shi 2012; Di Giammartino et al. 2011; Tian and Manley 2013; Elkon et al. 2013). It is estimated that over half of the eukaryotic genes produce alternatively polyadenylated transcripts (Shi 2012). Unlike alternative splicing, which has expanded dramatically during eukaryotic evolution, the prevalence of APA in yeast is comparable to that in metazoans (Ozsolak et al. 2010; Derti et al. 2012). APA isoforms may encode different proteins and/or have different 3′ untranslated regions (UTRs). In yeast, many APA events involve PAS found within the coding sequences (CDS) as well (Sparks and Dieckmann 1998; Yoon and Brem 2010; Mayer and Dieckmann 1989, 1991).

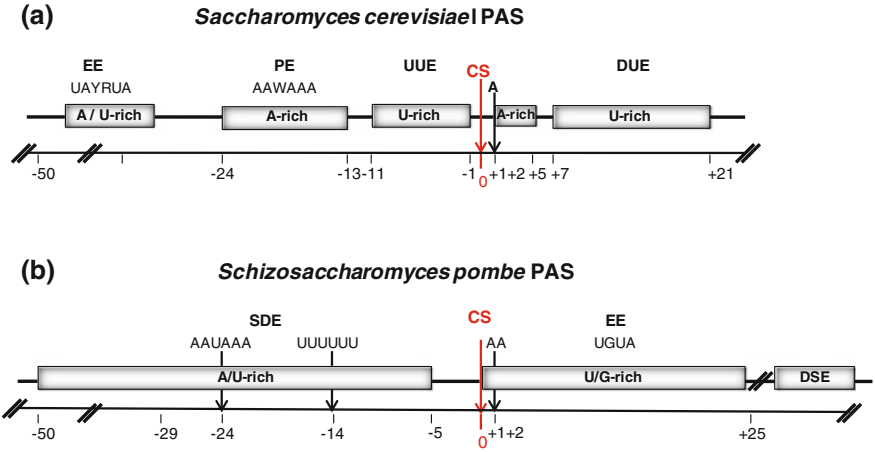


Fig. 3.1 Key cis-element for yeast poly(A) sites. The names, positions, consensus sequences of the known cis-element in *S. cerevisiae* (**a**) and *S. pombe* (**b**) PAS. *EE* Efficiency element, *PE* Positioning element, *UUE* Upstream U-rich element, *DUE* Downstream U-rich element, *SDE* Site determining element, *DSE* Downstream sequence element. See details in the text

Such APA isoforms are predicted to produce either truncated proteins or no protein product. Thus, APA significantly expands the proteome diversity and mRNA regulatory potential. The global APA profile is highly dynamic and regulated during development and in response to environmental cues (Shepard et al. 2011; Sandberg et al. 2008; Flavell et al. 2008; Graber et al. 2013). Aberrant APA regulation has also been implicated in a number of diseases, including cancer (Mayr and Bartel 2009; Jenal et al. 2012; Shi 2012; Di Giammartino et al. 2011).

The 3′ ends of mRNAs can be further processed in the cytoplasm (Richter 1999). For example, poly(A) tails can be extended or trimmed (Richter 1999). Other nucleotides, such as uracil, can be added (Rissland et al. 2007). These modifications play important roles in regulating the stabilities of target mRNAs (Scott and Norbury 2013). But for this article, we will focus on the 3′ end processing in the nucleus.

Fungal Poly(A) Site Sequence Features

Single gene studies and global analyses of *Saccharomyces cerevisiae* PAS have identified five key sequence elements (Fig. 3.1): (1) the A/U-rich Efficiency Element (EE) located at variable positions upstream of the cleavage site (CS) with the nucleotide consensus sequence UAYRUA (with Y: pyrimidine, and R: purine); (2) the A-rich Positioning Element (PE) located 10–30 nucleotides (nt) upstream of the CS. AAWAAA (W: A or U) is one of the most frequently found motifs in PE;

(3) the Upstream U-rich Element (UUE); (4) the CS; and (5) the Downstream U-rich Element (DUE) (Zhao et al. 1999). Recently, a short A-rich region from +2 to +5 nt (relative to the CS) has been suggested to influence the strength of the PAS (Fig. 3.1) (Moqtaderi et al. 2013). Functionally, the PE is most closely related to the AAUAAA hexamer in mammalian PAS (Zhao et al. 1999).

The *Schizosaccharomyces pombe* PAS have also been studied in some detail (Fig. 3.1) (Hansen et al. 1998; Birse et al. 1997; Humphrey et al. 1994). *S. pombe* PAS have an A-rich upstream sequence (called Site Determining Element (SDE)), equivalent of the aforementioned PE in *S. cerevisiae* PAS, and a UG-rich downstream element called Efficiency Element (EE) (Hansen et al. 1998; Birse et al. 1997; Humphrey et al. 1994). Genome-wide analyses suggest that AAUAAA is present in about 20 % of SDEs and UGUA is found in 24 % of EE in *S. pombe* PAS (Mata 2013; Schlackow et al. 2013; Ozsolak et al. 2010). Based on the high frequency of AAUAAA in SDEs, it has been suggested that *S. pombe* PAS are more similar to mammalian PAS (Humphrey et al. 1994; Chakraborty et al. 2002; Schlackow et al. 2013; Mata 2013). The distance between SDE and EE influences the efficiency of 3′ end formation, and the EE was proposed to enhance the binding of specific factors to SDE (Humphrey et al. 1994). An additional element located further downstream of the PAS, called the downstream element (DSE), also plays an important role in the transcription termination, most likely by inducing RNA pol II pausing (Birse et al. 1997; Hansen et al. 1998).

The specificity and activity of mRNA 3′ processing machineries in different yeast species seem quite similar. For example, *S. cerevisiae* PAS can be correctly processed in *S. pombe* and vice versa (Humphrey et al. 1991). A global comparison of mRNA polyadenylation in three yeast species, *S. cerevisiae*, *Kluyveromyces lactis*, and *Debaryomyces hansenii*, revealed several similarities (Moqtaderi et al. 2013). First, PAS in all these species share highly similar nucleotide composition and motifs (Fig. 3.1). Second, the position of the CS is highly heterogeneous. There are on average over 60 distinct CS located with a ~200 nt "end zone" for each PAS. In contrast, CS in mammalian PAS tend to cluster within 40 nt (Lee et al. 2007). Thus, the high heterogeneity of CS may be a widespread feature of fungal mRNA 3′ processing. Third, secondary structures may play an important role in determining PAS strength. RNA folding analyses predicted that the predominant PAS in these yeast species adopt a common configuration characterized by a double-stranded stem with the CS adjacent to a single-stranded domain. The weaker PAS seem less associated with such structures (Moqtaderi et al. 2013). On the other hand, species-specific differences in mRNA 3′ processing machinery also exist. For example, when large chromosomal fragments from *Debaryomyces hansenii* were introduced into *S. cerevisiae*, the polyadenylation within this region, especially the distribution of the CS within the end zone, adopted a pattern similar to that of the host strain. This observation suggests that the mRNA 3′ processing factors contribute to the species-specific polyadenylation profiles (Moqtaderi et al. 2013).

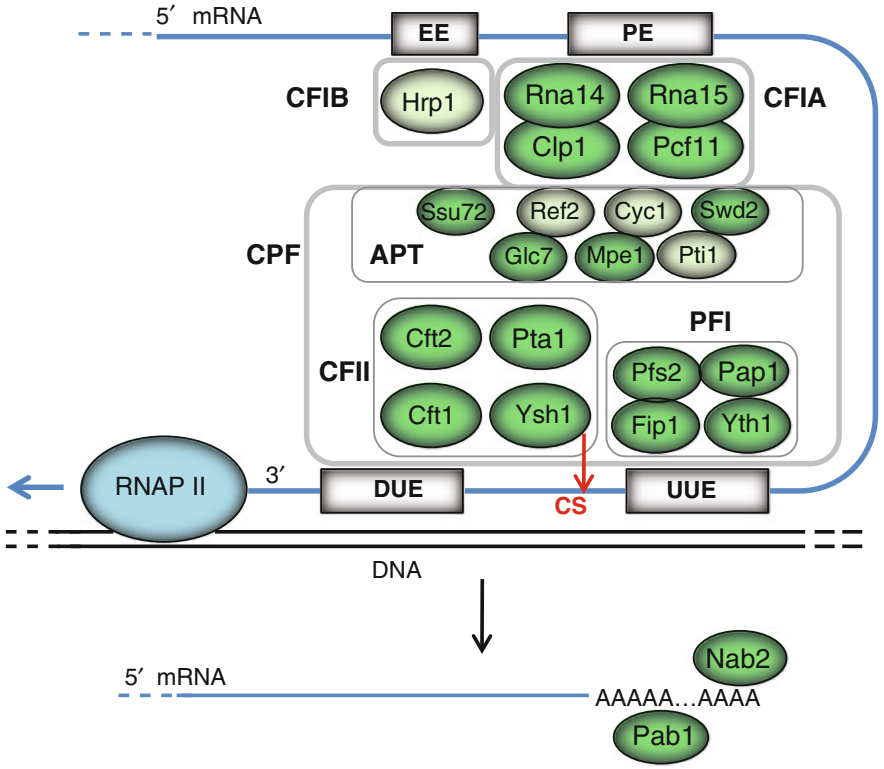


Fig. 3.2 The mRNA 3′ processing machinery in *Saccharomyces cerevisiae*. DNA templates are shown as *black lines* and RNA as a *blue line*. The mRNA 3′ processing factors conserved between yeast and mammals are marked as *dark green circles*. Subunits of complexes and subcomplexes are enclosed in *gray circles*

Fungal mRNA 3′ Processing Factors

There are over 20 known mRNA 3′ processing factors in *S. cerevisiae* (Fig. 3.2) (Zhao et al. 1999; Proudfoot 2004). At least 13 of them have been conserved up to mammals (Table 3.1). Interestingly, although *S. cerevisiae* and *S. pombe* do not express homologues of the mammalian Cleavage Factor I complex (CFIm), a putative homologue of the CFIm subunit CFIm25 can be found in filamentous fungus *Aspergillus oryzae*, and in the plant pathogens *Ustilago maydis* and *Magnaporthe oryzae* (Munsterkotter and Steinberg 2007; Franceschetti et al. 2011). There are no clear orthologs of CFIm68 in *M. oryzae* based on primary amino acid sequence. However, the *M. oryzae* protein Rbp35 interacts with the CFIm25 homologue in vivo (Franceschetti et al. 2011), and its RRM domain shares structural similarity with the metazoan CFIm68 RRM (Yang et al. 2010).

Table 3.1 Major 3′ processing factors in *S. cerevisiae* and *S. pombe*

Saccharomyces cerevisiae factor	*Schizosaccharomyces pombe* homologue	Domains	Functions	Mamalian homologue
CFIA				
Rna14	Rna14	HAT (Half A Tetratricopeptide repeat) : protein-protein interactions	Required for the specificity of Rna15 recognition of the PE	CstF77
Rna15	Ctf1	RRM : RNA binding	Binding to the PE	CstF64
Pcf11	Pcf11	CID: CTD interaction; Zinc fingers	Interaction with RNA PolI CTD. Transcription termination	Pcf11
Clp1		Walker A : ATP binding	CFIA Scaffolding; CPF /CFIA bridging	Clp1
CFIB				
Nab4/Hrp1		Two RRMs	Binding to EE. Important for NMD, APA, Copper stress response	
CFII				
Cft1/Yhh1	Cft1	Beta-Propeller repeats	Binding at the CS	Cpsf160
Cft2/Ydh1	Cft2	Metallo-Beta-Lactamase domain; Beta-CASP domain	Binding at the CS	Cpsf100
Ysh1/Brr5	Ysh1	Metallo-Beta-Lactamase domain; Beta-CASP domain	Putative endoribonuclease	Cpsf73
Pta1	Pta1		APT /CFII bridging	Symplekin
PFI				
Yth1	Yth1	CCCH zinc finger domain	Binds U-rich elements surrounding CS	Cpsf30
Pfs2	Pfs2	WD-40 repeat : protein-protein interaction	PFI /CFIA /CFII /APT bridging	Wdr33
Fip1	Fip1		Potential regulator of Pap1	Fip1

(continued)

Table 3.1 (continued)

Saccharomyces cerevisiae factor	*Schizosaccharomyces pombe* homologue	Domains	Functions	Mamalian homologue
Other CPF subunits				
Pap1	Pla1		Catalyses adenosine addition to the 3′ cleavage product	PAP
Glc7	Dis2; Sds21		Type 1 protein phosphatase	PP1
Mpe1	SPBP8B7.15c	Zinc knuckle		Rbbp6
Ssu72	Ssu72		Phosphatase	Ssu72
Swd2	Swd2.2 (SPAC824.04) : CPF; Swd2 : Set1 complex	WD40 repeat	Links 3′ processing with chromatin modifications? Increases weak PAS usage	Wdr82 : Set1 complex
Ref2				
Pti1		RRM	snoRNA 3′ formation	
Syc1			Negative regulator?	

These structural and functional similarities raise the possibility that *RBP35* is the functional ortholog of CFIm68 in filamentous fungi. Below we discuss the structure and functions of the major fungal 3′ processing factors in detail.

mRNA 3′ Processing Factors in S. cerevisiae

Most mRNA 3′ processing factors in *S. cerevisiae* are essential for viability and they form three major complexes: *Cleavage Factor IA (CFIA), Cleavage Factor IB (CFIB), and Cleavage and Polyadenylation Factor (CPF)* (Zhao et al. 1999). CPF consists of three subcomplexes: Cleavage Factor II (*CFII*), Polyadenylation Factor I (*PFI*), and Associated with Pta1 (*APT*). Additionally, the C-terminal domain (*CTD*) of RNA polymerase II (RNA pol II) binds to many mRNA 3′ processing factors and facilitates the recruitment of these factors to nascent RNAs in a co-transcriptional manner (see also Chap. 1).

CFIA

The CFIA complex contains four subunits, Rna14, Rna15, Pcf11, and Clp1. *RNA14* and *RNA15* are the homologs of the mammalian CstF77 and CstF64, respectively. *S. cerevisiae* does not appear to encode a homolog of the third mammalian CstF subunit, CstF50 (Zhao et al. 1999). Depletion of Rna14 or Rna15 leads to global shortening of poly(A) tails and a defect in PAS recognition (Minvielle-Sebastia et al. 1994; Mandart and Parker 1995). Similar to its mammalian homolog CstF77, Rna14 seems to serve as a scaffold through interactions with Rna15, CFIB (see below), and RNA POL II CTD (Noble et al. 2004). Also similar to its mammalian homolog CstF64, Rna15 contributes to PAS recognition by directly binding to RNA via its RRM domain (Gross and Moore 2001). However, Rna15 binds to the A-rich PE upstream of the CS while CstF64 binds specifically to the U/GU-rich regions downstream of CS (MacDonald et al. 1994; Takagaki and Manley 1997; Yao et al. 2012). Recognition of the PE by Rna15 also requires CFIB (Hrp1/Nab4) (Leeper et al. 2010).

PCF11 and *CLP1* are also conserved from yeast to human (Darmon and Lutz 2012). Pcf11 interacts with the RNA pol II CTD through its N-terminal CTD Interacting Domain (CID) (Meinhart and Cramer 2004; Barilla et al. 2001; Licatalosi et al. 2002). RNA pol II CTD is a unique protein domain that consists of 26 (yeast) to 52 (human) highly conserved heptapeptide repeats of the consensus sequence: YSPTSPS (Buratowski 2003; Hirose and Manley 2000; Bentley 2005). The CTD is highly phosphorylated and its phosphorylation is tightly regulated during the transcription cycle. For example, Ser5 is phosphorylated early in transcription and helps to recruit the capping enzymes. On the other hand, Ser2 phosphorylation is low at the promoter regions, but accumulates during transcription

elongation and peaks near the 3′ ends of genes and is important for recruiting mRNA 3′ processing factors (Buratowski 2003; Hirose and Manley 2000; Bentley 2005). For example, Pcf11 CID specifically interacts with CTD phosphorylated at Ser2 (Licatalosi et al. 2002; Meinhart and Cramer 2004). In addition to its role in mRNA 3′ processing, Pcf11 functions in transcription termination by bridging the RNA pol II CTD to the nascent transcripts and dismantling the transcription elongation complex (Zhang et al. 2007; Zhang and Gilmour 2006).

Clp1 interacts with Pcf11 and the CFII subunits Ysh1/Brr5, thereby linking CFIA to CPF (Minvielle-Sebastia et al. 1997; Kessler et al. 1996). Clp1 contains a Walker A motif, a known ATP/GTP-binding domain, but no ATPase activity has been detected (Noble et al. 2007). The human Clp1 protein has been shown to possess RNA-specific 5′-OH polynucleotide kinase activity (Weitzer and Martinez 2007). However, yeast Clp1 seems to lack this activity and mutations in the kinase domain do not affect viability (Ramirez et al. 2008), indicating that the RNA kinase activity is not required for mRNA 3′ processing in yeast.

CFIB

CFIB is composed of one unique subunit, *Hrp1/Nab4* (16, 62, 160). Although Hrp1/Nab4 has no sequence homology with any mammal protein, it may share similar structures with the mammal splicing factor hnRNP A1 (Kessler et al. 1997). *HRP1/NAB4* is essential for cell viability and is required for cleavage and polyadenylation. Hrp1/Nab4 directly binds to the U-rich EE via its two RRM domains (Chen and Hyman 1998; Perez-Canadillas 2006). Its depletion in vivo leads to a global decrease in poly(A) tail length (16, 62). It interacts with Rna14 and Rna15, and shuttles between the nucleus and the cytoplasm (Kessler et al. 1997). Moreover, Hrp1/Nab4 was shown to regulate APA and stress response (further discussed in the APA section) (Kim Guisbert et al. 2007).

CPF

CPF contains homologues of all the major subunits of mammal Cleavage and Polyadenylation Specificity Factor (CPSF) and they comprise three subcomplexes: CFII, PFI, and APT (Zhao et al. 1999).

The CFII subcomplex. CFII contains four proteins: Cft1/Yhh1 (Cleavage Factor Two 1), Cft2/Ydh1 (Cleavage Factor Two 2), Ysh1/Brr5 (Yeast 73 kDa Homolog 1), and Pta1 (Pre-Trna Accumulation 1). They are homologous to the mammalian CPSF160, CPSF100, CPSF73, and symplekin, respectively (Zhao et al. 1999; Darmon and Lutz 2012). The CFII subunits are functionally similar to their mammalian counterparts as well. Cft1/Yhh1 binds to the mRNAs in the vicinity of the CS via a Beta-propeller repeat domain (Fig. 3.2) (Stumpf and Domdey 1996; Dichtl

et al. 2002b). Moreover, Cft1/Yhh1 interacts with RNA pol II CTD and is essential for mRNA transcription termination (Dichtl et al. 2002b). Similar to CPSF 100 and CPSF73, Cft2/Ydh1 and Ysh1/Brr5 both contain a putative metallo-beta-lactamase domain and a beta-CASP domain. However, Cft2/Ydh1 is not able to bind metal ions (Mandel et al. 2006). Cft2/Ydh1 binds the mRNA at a region encompassing the CS (Zhao et al. 1997). Cft2/Ydh1 interacts with other CFII subunits, Pfs2 (PFI subunit), Ssu72 (APT subunit), and the RNA pol II CTD (Kyburz 2003). Similar to CPSF73, Ysh1/Brr5 is believed to be the endonuclease for mRNA 3′ processing in budding yeast. Consistent with this conclusion, mutations in Ysh1/Brr5 that disrupt zinc binding are lethal (Mandel et al. 2006; Ryan et al. 2004). Pta1 shares homology with Symplekin, a scaffolding factor in the mammalian CPSF complex (Takagaki and Manley 2000). Pta1 is believed to bridge the APT complex with CFII complex through multiple interactions (Nedea et al. 2008).

The CPF: PFI complex. The PFI subcomplex contains Yth1 (Yeast 30 kDa Homolog 1), Pfs2 (Polyadenylation Factor Subunit 2), Fip1 (Factor Interacting with Poly(A) polymerase 1), Mpe1 (Mutant PCF11 Extragenic suppressor 1), and Pap1 (Zhao et al. 1999).

Yth1 is related to the mammal RNA-binding zinc finger protein CPSF30 (64 % similarity and is essential for in vitro cleavage and polyadenylation (Barabino et al. 1997). Yth1 contains five CCCH zinc finger domains, and the second one has been shown to be critical for mRNA 3′ processing. Yth1 binds to the U-rich element surrounding the CS: the UUE and DUE sequences (Barabino et al. 1997). Yth1 interacts with Fip1 and Ysh1/Brr5, subunit of CFII subcomplex (Barabino et al. 1997; Tacahashi et al. 2003; Helmling et al. 2001).

Fip1 is an intrinsically disordered/unstructured protein that shares 52 % similarity with mammal RNA-binding protein Fip1 (Meinke et al. 2008; Darmon and Lutz 2012). Although Fip1 is not required for cleavage, elimination of the C-terminal half leads to a general shortening of poly(A) tail in vivo (Preker et al. 1995). Fip1 interaction with Pap1 was suggested to regulate Pap1 poly(A) polymerase activity (Preker et al. 1995; Helmling et al. 2001). Moreover, as mutations specifically disrupting Fip1–Pap1 interactions are lethal, it is likely that Fip1 mediates the recruitment of Pap1 to the PAS (Helmling et al. 2001). Fip1 also interacts with Rna14 and Pfs2 (Ohnacker et al. 2000).

Pfs2 is a WD-40 repeat protein and the homolog of the mammal WD40 repeat protein Wdr33 (58 % similarity) (Darmon and Lutz 2012). Pfs2 is required for cleavage and polyadenylation. Pfs2 links PFI with CFIA, CFII, and APT subcomplexes through its interaction with Fip1, Rna14, Ysh1/Brr5, and Swd2 (Ohnacker et al. 2000).

Pap1 is required for polyadenylation but not for cleavage in vitro (Lingner et al. 1991). Structural and enzymatic properties of Pap1 are highly conserved in fungi, as exemplified by studies of *Candida albicans* and *S. pombe* poly(A) polymerase (Bougie and Bisaillon 2007). Pap1 is posttranslationally modified by phosphorylation and ubiquitylation during the cell cycle. Phosphorylation of Pap1 occurs during S and G2 phases, and this modification inhibits Pap1 activity (Mizrahi and Moore 2000). Several proteins, including Fip1, Cft1, and Pta1, interact with Pap1

and could potentially modulate its activity (Nedea et al. 2003; Ezeokonkwo et al. 2012). Pap1 also interacts with the RNA-binding protein Nab6, which was suggested to bind poly(A) mRNA to increase their stability, and to target more specifically mRNAs encoding for proteins of the cell wall (Ezeokonkwo et al. 2012). Pap1 also influences PAS choice (Mandart and Parker 1995).

The CPF: APT subcomplex. CPF contains additional factors that form a third subcomplex called the APT complex (Associated with Pta1) (Nedea et al. 2003). Some of these factors have homologs in metazoan, including Glc7 (GLyCogen 7, homolog of the mammalian phosphatase PP1), Ssu72, and Swd2 (Darmon and Lutz 2012). The other APT component appear to be specific to yeast, including Ref2 (RNA End Formation 2), Pti1 (PTa1p Interacting protein), and Syc1 (Similar to Ysh1 C-terminal 1) (Darmon and Lutz 2012).

Ssu72 is required for cleavage but not for polyadenylation (He et al. 2003). Ssu72 is a protein phosphatase that specifically dephosphorylates RNA pol II CTD at Ser5 (Krishnamurthy et al. 2004; Hausmann et al. 2005). It has been suggested that Ssu72 functions to regenerate hypophosphorylated RNA pol II for new rounds of transcription. However, Ssu72 phosphatase activity seems not required for mRNA 3′ processing itself (Krishnamurthy et al. 2004; Hausmann et al. 2005). Besides Pta1, Ssu72 interacts with Cft2/Ydh1 and with RNA pol II subunit, Rpb2 (Krishnamurthy et al. 2004; Hausmann et al. 2005; Dichtl et al. 2002a). Even though the *Kluyveromyces lactis* Ssu72 protein shares 76 % identity with its *S. cerevisiae* counterpart, it cannot functionally complement *S. cerevisiae* Ssu72, indicating functional divergence during fungal evolution (Rodriguez-Torres et al. 2013).

Glc7 is the homolog of the mammalian protein phosphatase PP1 (Darmon and Lutz 2012). Glc7 is specifically required for cleavage, but not for polyadenylation (He and Moore 2005). Pta1 was identified as the Glc7 substrate in the mRNA 3′ processing machinery. It has been proposed that Pta1 goes through a phosphorylation-dephosphorylation cycle during mRNA 3′ processing and Glc7-mediated dephosphorylation is essential for the transition between cleavage and polyadenylation (He and Moore 2005). Glc7 is regulated by another APF subunit, Ref2 (RNA end formation 2) (Nedea et al. 2008). Ref2 directly binds to RNA and is required for the efficient processing at weak poly(A) sites (Russnak et al. 1995). Additionally, Ref2 mediates Glc7 association with the CPF complex. In the absence of Ref2, Glc7 dissociates from CPF, which results in defects in transcription termination at snoRNA genes (Nedea et al. 2008).

Swd2 is a WD-40 repeat protein essential for cell viability. It was first identified as part of Set1 (SET (Su(var)3-9; Enhancer of zeste; Trithorax) domain containing 1) /COMPAS (Complex Proteins Associated with Set1), which is essential for histone H3 methylation at lysine 4 (H3K4me) (Roguev et al. 2001; Miller et al. 2001). This epigenetic mark is important for gene expression and also for rDNA and telomeric heterochromatin silencing (Eissenberg and Shilatifard 2010). The presence of this epigenetic mark necessitates Swd2 (Cheng et al. 2004). Swd2 was also identified as part of the APT complex and functionally interacts with Ref2 (Nedea et al. 2003). Even though Swd2 is not essential for in vitro cleavage and polyadenylation, it is

required for RNA pol II transcription termination (Cheng et al. 2004). However, these two functions of Swd2 do not seem to be tightly coupled (Cheng et al. 2004).

Mpe1 has a putative zinc knuckle domain that may mediate RNA interactions and plays critical roles in mRNA 3′ processing by promoting the specific interactions between CPF and the pre-mRNAs (Vo et al. 2001). Mpe1 shares limited homology with the mammalian protein Rbbp6, which interacts with Rb and p53 and has been implicated in cancer (Shi et al. 2009; Pugh et al. 2006; Sakai et al. 1995). The functions of Mpe1 in mRNA 3′ processing remain poorly understood.

Syc1 shares homology with the C-terminal domain of Ysh1/Brr5 (Zhelkovsky et al. 2006). As mentioned above, *YSH1/BRR5* is essential for cell viability and mRNA 3′ processing. *SYC1* is not essential, but its deletion rescues the growth and mRNA 3′ processing defects in ysh1/brr5 mutant, indicating that Syc1 is a negative regulator of mRNA 3′ processing (Zhelkovsky et al. 2006).

Pti1 shares homology with Rna15 and the mammalian CstF64 and interacts with Pta1, but Pti1functions in mRNA 3′ processing have not been characterized in detail (Qu et al. 2007).

Poly(A) Binding Proteins

Poly(A) **B**inding **P**roteins (Pabps) play important roles in poly(A) tail length control (Mangus et al. 2003). *S. cerevisiae* encodes two main Pabps, Pab1 and Nab2, and they are homologous to the mammalian proteins PABPC1 and ZC3H14 (Soucek et al. 2012). Both proteins are essential for cell viability and depletion of either proteins leads to a global lengthening of poly(A) tails in vivo (Sachs and Davis 1989; Anderson et al. 1993). Nab2 is believed to be the major Pabp in the nucleus. It is co-transcriptional recruited to the nascent transcripts (Soucek et al. 2012). Nab2 physically interacts with Hrp1 and genetically interacts with Pap1, Rna15, and Syc1 (Soucek et al. 2012; Yu et al. 2008; Kerr et al. 2013). Nab2 interacts with RNAs through its zinc finger domains (Anderson et al. 1993; Marfatia et al. 2003). However, it remains poorly understood how Nab2 contributes to poly(A) tail length control. Additionally, Nab2 interacts with the Mlp1 (Myosin Like Protein 1), a factor involved in the nuclear retention of unspliced mRNAs and the nuclear exosome subunit Rrp6 (Green et al. 2003). These observations suggest that Nab2 contributes to mRNA quality control by targeting misprocessed RNAs to the exosome for degradation (Schmid et al. 2012; Soucek et al. 2012). Following mRNA export, Nab2 is believed to be replaced by Pab1 during the mRNP remodeling (Soucek et al. 2012). Pab1 contains four RRM domains and is associated with CFIA through the interaction with Rna15 (Amrani et al. 1997). In addition to its role in nuclear mRNA 3′ processing, Pab1 also mediates poly(A) shortening to promote translation in vivo (Sachs and Davis 1989). Pab1 recruits the Pab1-dependent Poly(A) Nuclease (PAN) to trim the poly(A) tails (Mangus et al. 2004). Because overexpression of Pab1 cannot rescue the hyperadenylation defect in nab2-deficient cells, these proteins have nonoverlapping functions (Hector et al.

2002). For example, Pab1 is able to bind mRNA with poly(A) tails as short as 10 nt, whereas Nab2 associates mainly with mature poly(A) tails (60–80 nt) (Hector et al. 2002).

S. pombe 3′ Processing Factors

Despite the fact that *S. pombe* and *S. cerevisiae* are evolutionary rather distant, their mRNA 3′ processing machineries are more similar to each other than to the mammalian system. For example, the *S. pombe* poly(A) polymerase Pla1 shows a higher sequence similarity with *S. cerevisiae* Pap1 (55 % identity) than with its mammalian homologues (88). Consistently, *S. cerevisiae* Pap1 can be functionally substituted both in vivo and in vitro by Pla1, whereas Pla1 cannot replace mammalian Pap1 in vitro (Ohnacker et al. 1996). These results suggest that Pla1 is able to interact with *S. cerevisiae* mRNA 3′ processing factors efficiently enough to correctly process pre-mRNAs. Due to these similarities, the *S. pombe* ura4 transcript can be cleaved and polyadenylated in vitro in *S. cerevisiae* cell extracts, and conversely, the *S. cerevisiae* cyc1 mRNA is correctly processed in vivo when expressed in *S. pombe* (Humphrey et al. 1991).

On the other hand, some of 3′ processing factors in *S. pombe* seem functionally closer to their mammalian homologs than to their budding yeast counterparts. For example, *S. pombe* Ctf1 (also called spCstF-64) is homolog to *S. cerevisiae* Rna15 and mammalian CstF64 (Aranda and Proudfoot 2001). Unlike Rna15, which recognizes the PE upstream of CS (39), Ctf1 binds to the EE downstream to the CS (Dichtl and Keller 2001). This is similar to its mammalian homolog CstF64, which has been shown to bind specifically to U/GU-rich sequences downstream of the CS (MacDonald et al. 1994; Takagaki and Manley 1997; Yao et al. 2012)(Table 3.1).

Interestingly, mRNA 3′ processing defects have been shown to manifest in some unexpected phenotypes in *S. pombe*. For example, mutations in *PFS2* gene cause chromosome segregation defects, which are believed to be downstream effect of mRNA 3′ processing and transcription termination malfunctions (Wang et al. 2005). Supporting this conclusion, transcription termination defects are observed in cells deficient for Pfs2 and the chromosome segregation defects are suppressed by overexpression of another mRNA 3′ processing factor Cft1(Wang et al. 2005). In keeping with the link between transcription termination and cell cycle , Dhp1, a 5′–3′ exonuclease homologous to Rat1 in *S. cerevisiae*, is also required for chromosome segregation (Shobuike et al. 2001; Sugano et al. 1994).

Although most mRNA 3′ processing factors in *S. pombe* have not been studied in detail, the Pabps have been characterized. *S. pombe* encodes for two RRM-containing Pabps, Pabp and Pab2 (Perreault et al. 2007; Thakurta et al. 2002). These proteins are nonessential for cell viability, suggesting functional redundancy. Pabp is the homolog of *S. cerevisiae* Pab1, and has been shown to be involved in mRNA export (Thakurta et al. 2002). Pab2 shares 47 % identity and 66 % similarity with human PABPN1, and possess a coiled-coil region, an RRM,

and a C-terminal arginine-rich domain. Pab2 binds to RNA poly(A) tails in vitro, and regulates mRNA poly(A) length in vivo (Perreault et al. 2007). Indeed, the maximum length of the poly(A) tail exceeds 226 nt in pab2-deficient cells as compared to 120 nt in wild type cells. Pab2 has been shown to self-associate in an RNA-independent manner. Similar to PABPN1, Pab2 is methylated at the R residues of the R-rich domain by Rmt1, a type I protein arginine N-methyltransferase (Perreault et al. 2007). R methylation is important for Pab2 oligomerization, but not for its nuclear localization or its function in regulation of poly(A) tail length. Mutations in the human *PABPN1* gene are linked to oculopharyngeal muscular dystrophy (OMPD), a disease characterized by fibrous inclusions in the nuclei of skeletal muscle fibers (Jenal et al. 2012; de Klerk et al. 2012). Interestingly, Pab2 overexpression leads to growth defects mediated by the R-rich domain, exacerbated when rmt1 is deleted, suggesting that elevated levels of unmethylated Pab2 is toxic for the cells (Perreault et al. 2007). Altogether these data suggest that Pab2 is functionally similar to its human homolog. Chromatin immunoprecipitation data indicate that although Pab2 binds to the poly(A) tail of mRNA, it might be recruited at earlier steps of transcription through its interaction with the large RNA pol II subunit Rpb1 (Perreault et al. 2007).

Regulators of mRNA 3′ Processing

Several factors have been identified as regulators of mRNA 3′ processing. One of the negative regulators is the RNA-binding protein Npl3 (Nuclear Protein Localization 3) (Bucheli and Buratowski 2005). Npl3 contains two RRMs and a domain that is rich in Serine/Arginine (SR) dipeptide repeats, a domain structure that is similar to the SR family splicing regulators in higher eukaryotes (Graveley 2000). Npl3 is required for the correct splicing of several mRNAs by mediating the recruitment of splicing factors via direct interaction (Kress et al. 2008). In addition to its functions in splicing, Npl3 stimulates transcription elongation through interactions with RNA pol II and impedes efficient transcription termination by competing with Rna15 for binding to the A-rich PE of the PAS (Dermody et al. 2008; Deka et al. 2008). Phosphorylation of Npl3 by Casein Kinase 2 (CK2) decreases its interaction with the RNA pol II and its ability to bind RNA, thereby promoting transcription termination (Dermody et al. 2008). Moreover, Npl3 phosphorylation stimulates a negative autoregulation by promoting the distal PAS usage of *NPL3* transcript, which leads to a decrease in Npl3 protein level (Lund et al. 2008).

The mRNA export adaptor Yra1 has also been shown to regulate mRNA 3′ processing. Yra1 negatively regulates 3′ end formation by competing with Clp1 for interaction with Pcf11, and depletion of Yra1 leads to changes in the global APA profile (Johnson et al. 2009, 2011).

Factors Required for S. cerevisiae Histones mRNA 3′ Processing

The expression of replication-dependent histone genes is highly regulated during the cell cycle to allow accumulation of histone mRNAs specifically during the S phase (Marzluff et al. 2008). In metazoans, the 3′ ends of replication-dependent histone mRNAs are formed by an endonucleolytic cleavage step without polyadenylation (Marzluff et al. 2008). This process involves the recognition of a highly conserved stem-loop by the SLBP protein and the downstream sequences by the U7 snRNP at the 3′ ends of the mRNAs. But mRNA 3′ processing factors, such as CPSF and CstF, are also required. The metazoan histone mRNA 3′ processing activities are regulated in a cell cycle-specific manner (Marzluff et al. 2008).

In contrast to metazoans, the 3′ ends of histone mRNAs in fungi, plants, and protozoa generated through the regular cleavage/polyadenylation mechanism and the 3′ processing of these mRNA requires the canonical mRNA 3′ processing factors, including Rna14, Pcf11, Rna15, and Pap1 (Fahrner et al. 1980; Canavan and Bond 2007). However, recent studies have implicated Sen1, a putative helicase required for the 3′ processing of many nonpolyadenylated RNAs, in yeast histone mRNA 3′ processing (Beggs et al. 2012). Additionally, in *S. cerevisiae*, the poly(A) tails of histone mRNAs, which are 20–50 nt, are shorter than the average length (70–90 nt), and their poly(A) tails shorten during the *S* phase (Beggs et al. 2012). *S* phase-specific inhibition of Pap1 activity by phosphorylation might be a potential mechanism for histone mRNA poly(A) tail length control. Further studies are needed to understand how histone mRNAs levels are regulated by the cell cycle and how such mechanisms evolved during evolution.

Release of mRNA 3′ Processing Factors After Polyadenylation

Although the assembly of the mRNA 3′ processing machinery has been studied extensively, how these factors are released following polyadenylation remains poorly understood. An important insight came from the observation that mutations in factors involved in mRNA export or the assembly of export-competent mRNPs (such as Mex67), lead to a defect in the release of mRNA 3′ processing factors from mRNAs (Qu et al. 2009). This suggest that a remodeling of mRNPs takes place during which mRNA export factors may replace 3′ processing factors on polyadenylated mRNAs (Qu et al. 2009). This functional coupling may help to ensure that only fully processed mRNAs are targeted for export, but the mechanistic details of this mRNP remodeling step remain unclear.

Alternative Polyadenylation and its Regulation in Fungi

Recent global studies have revealed that APA is surprisingly widespread in yeast. For example, it is estimated that 40–70 % of the *S. cerevisiae* and *S. pombe* genes produce alternatively polyadenylated mRNAs whose CS are separated by 50 nt or more (Ozsolak et al. 2010; Mata 2013; Schlackow et al. 2013; Moqtaderi et al. 2013; Yoon and Brem 2010). However, as mentioned earlier, there seems to be a high level of heterogeneity in the position of the CS in yeast (Moqtaderi et al. 2013). Rather than one or a few distinct CS, cleavage/polyadenylation occurs in a ~200 nt "end zone." Thus, it may be difficult to distinguish between CS heterogeneity from the same PAS and APA. However, recent studies clearly demonstrated that APA is widespread in fungi (Ozsolak et al. 2010; Mata 2013; Schlackow et al. 2013; Moqtaderi et al. 2013; Yoon and Brem 2010). In *S. cerevisiae*, more than 600 genes use PAS within the CDS, producing truncated transcripts. Interestingly, a motif, GAAGAAGA, is enriched in the 50 nucleotides upstream of the intragenic CS. These truncated transcripts originate mainly from genes involved in stress response and meiosis (Yoon and Brem 2010). Indeed, APA has been implicated in cellular responses to many types of stress and in the regulation of meiotic gene expression, which are discussed in details below.

APA Regulation and Metabolism

One of the first examples of APA regulation in budding yeast was described for the gene *CBP1* (Cytochrome b processing 1) (Mayer and Dieckmann 1989). Cbp1 is required for the expression of the mitochondrial gene encoding Cytochrome B, a component of the electron transport chain in respiration. In fermenting cells, two *CBP1* APA isoforms are produced (Mayer and Dieckmann 1989). The short isoform uses a PAS within the CDS and the resulting truncated mRNA does not code for any protein product. The longer isoform encodes the functional Cbp1 protein. Following induction of respiration by switching to a nonfermentable carbon source, there is a shift in *CBP1* APA pattern from the long to the short isoform while the total mRNA level remains unchanged (Sparks and Dieckmann 1998; Mayer and Dieckmann 1989). Three additional mRNAs were later shown to undergo the same type of respiration-dependent APA change: *AEP2/ATP13*, which is also necessary for respiration, and *RNA14* and *SIR1*. Like *CBP1*, *AEP2* encodes a factor required for the expression of a mitochondrial respiration gene (*ATP9*). Rna14 is a mRNA 3′ processing factor (see "CFIA"), and *SIR1* encodes a mating type locus silencing factor. Suppression of the short *CBP1* transcript leads to the constitutive production of elevated levels of the long *CBP1* transcript independently of respiration induction, which in turn results in the accumulation of the mitochondrial *CYTOCHROME B* mRNAs (Sparks and Dieckmann 1998). These

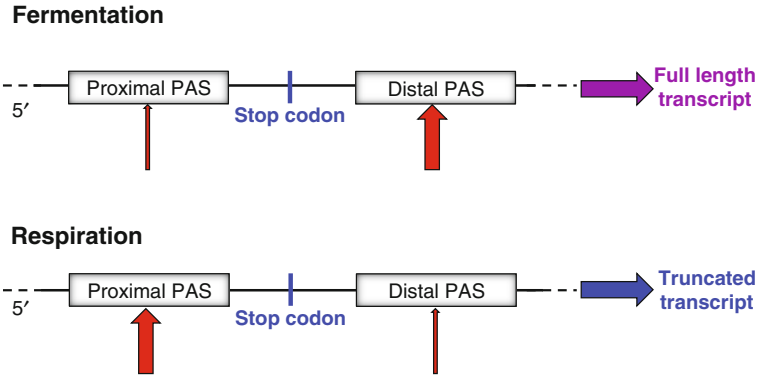


Fig. 3.3 Metabolism-mediated APA regulation. In fermenting *Saccharomyces cerevisiae* cells, polyadenylation occurs mainly at the distal poly(A) site downstream of the coding sequences, leading to production of the full-length mRNAs. In respiring cells, polyadenylation shifts to the proximal poly(A) sites in the coding sequences, leading to the production of truncated transcripts

observations indicate that the APA may be a mechanism for respiration-dependent regulation of gene expression (Fig. 3.3). However, the mechanism and biological significance of APA regulation by metabolism remain to be determined.

APA Regulation and DNA Damage Response

APA changes have been observed in response to DNA damage (Fig. 3.4). For example, polyadenylation of *RPB2* (RNA Polymerase B 2) and *CBP1* mRNAs switches from the proximal to the distal PAS upon UV irradiation (Yu and Volkert 2013). A global study detected similar APA changes for over 2,000 genes under similar conditions (Graber et al. 2013). Two possible mechanisms have been proposed, which are not mutually exclusive. First, the transcription elongation rate has been suggested to play an important role (Yu and Volkert 2013). A pharmacologically induced decrease in transcription elongation rate abolishes the APA changes in *RPB2* mRNAs following DNA damage (Yu and Volkert 2013), indicating that fast transcription elongation rate promotes skipping of the proximal PAS. Second, UV-induced DNA damage has been shown to cause a reduction in the protein levels of CPF subunits and in turn lower mRNA 3′ processing activity. The decrease in the 3′ processing activity results in the preferential recognition of the distal PAS as they are intrinsically stronger and have higher affinity for the mRNA 3′ processing machinery (Graber et al. 2013) (Fig. 3.4). Interestingly, a transient inhibition of the mRNA 3′ processing machinery has been observed in mammalian cells (Kleiman and Manley 2001). This is mediated by sequestration of CstF50 by the BRCA1-BARD1 complex following DNA damage (Kleiman and

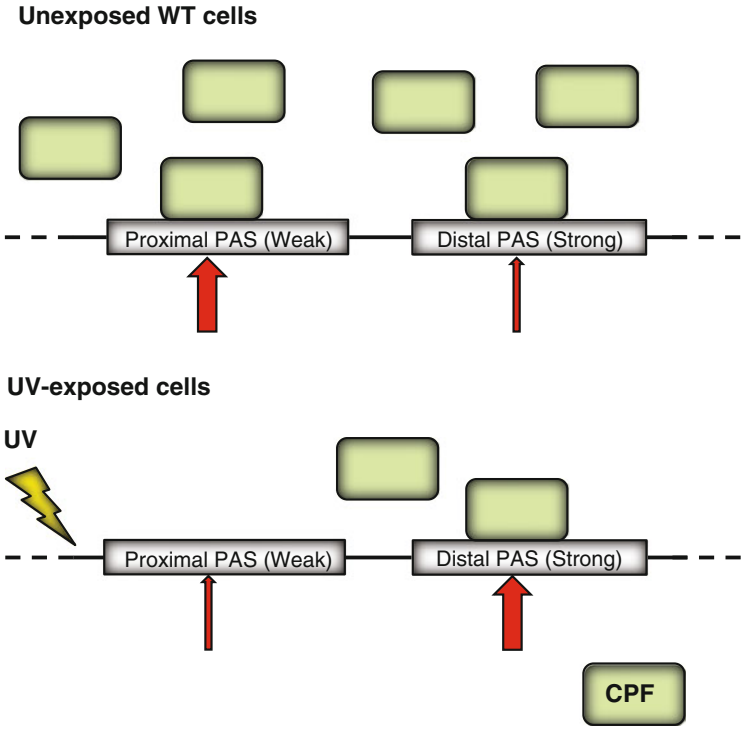


Fig. 3.4 APA regulation in DNA damage response. In unstressed cells, high levels of CPF allows for the recognition of weak poly(A) sites in the coding sequences, leading to the production of truncated transcripts. Following UV-induced DNA damage, CPF levels decrease and the remaining CPF preferentially binds to the strong distal poly(A) sites downstream of the coding sequences, leading to production of full-length mRNAs

Manley 2001). It will be of great interest to determine how CPF levels are regulated and what the functional importance of these APA changes is for cellular survival after DNA damage.

Nab4/Hrp1-Mediated APA Regulation and Copper Stress

The general 3′ processing factor Nab4/Hrp1 has been identified as an important APA regulator (Kim Guisbert et al. 2007). The *SUA7* gene (Suppressor of Upstream AUG 7) produces two APA isoforms. The long isoform is more abundant in exponentially growing cells while the short isoform accumulates during stationary phase (Hoopes et al. 2000). Nab4/Hrp1 was found to be critical for regulating the cell cycle-dependent ratio of the isoforms (Kim Guisbert et al. 2007). Nab4/Hrp1 binds to a UA-rich motif and promote the usage of adjacent

PAS. Additionally, the protein level of Nab4/Hrp1 is also important: high levels of Nab4/Hrp1 promote the usage of the proximal PAS in *SUA7* mRNA (Kim Guisbert et al. 2007). However, it has not been determined whether and how Nab4/Hrp1 protein level is regulated in cell cycle. Interestingly, nab4/hrp1 mutant stains are extremely resistant to high concentration of copper. This is due to the APA change in the *CTR2* gene, which encodes a copper transporter. In nab4/hrp1 mutant strain, the level of the *CTR2* mRNA isoforms with the longest 3′ UTR increases (Kim Guisbert et al. 2007). These observations suggest that APA regulation plays an important role in cellular stress induced by copper and perhaps other metals.

APA Regulation of Meiotic Gene Expression

A subset of meiotic genes in *S. pombe* was also shown to undergo APA upon meiosis induction, and it was suggested to be important for the regulation of their expression (McPheeters et al. 2009; Cremona et al. 2011; Potter et al. 2012). The mechanisms involved are discussed in a following paragraph.

The first example of meiosis-dependent APA regulation was demonstrated for the meiotic gene *CRS1* (McPheeters et al. 2009). *CRS1* mRNAs are polyadenylated at two PAS. In vegetative cells, *CRS1* mRNAs are actively degraded by the Mmi1 pathway as described later ("mRNA 3′ Processing in the Regulation of S. pombe Meiotic Genes"). Upon meiosis, Mmi1-mediated repression is alleviated and *CRS1* mRNA undergoes splicing-coupled polyadenylation at both proximal and distal PAS. Even though the ratio between the short and the long isoforms slightly changes during the time course of meiosis, the distal PAS is always more predominantly used over the proximal PAS (McPheeters et al. 2009; Chen et al. 2011). Later studies identified additional meiotic transcripts using meiosis specific 3′ processing-dependent regulation (Cremona et al. 2011; Potter et al. 2012). These transcripts were found to utilize more than one PAS upon meiosis induction, but again in this case the proximal PAS usage relatively to the distal PAS was not studied in detail. Although the biological consequences of meiotic-dependent APA are still not clear, APA might be an additional way to regulate the proper timing of activation of these genes during sexual differentiation progression.

Other Examples of Alternative Polyadenylation in Fungi

Kluyveromyces lactis (*K. lactis*) *CYC1* mRNA (*KLCYC1*) was shown to use two distinct PAS, whereas *S. cerevisiae CYC1* mRNA has only one PAS (Freire-Picos et al. 1995). *CYC1* encodes the iso-1-cytochrome c factor, and *KLCYC1* is essential for respiratory growth in *K. lactis*. Specifically the longer APA isoform has been suggested to be responsible for an increase biomass production during respiration, and an inhibition of ethanol production during fermentation of *K. lactis* (Seoane et al.

2005). Comparison of the sequences surrounding the two CS revealed a common feature. There is an AACAA motif a few nucleotides upstream of the CS, and an AU-rich region just upstream of the AACAA motif only for the proximal PAS. The ratio between the two isoforms changes according to cellular growth conditions: the distal PAS usage increases with the optical density of the culture (OD) when the proximal PAS usage remains constant. *KLCYC1* mRNA was correctly processed at the two PAS when transformed into *S. cerevisiae*, but the growth-dependent APA change was not observed, indicating the lack of a specific regulatory factor in *S. cerevisiae* (Freire-Picos et al. 2001). The mechanism for *KLCYC1* APA regulation remains unclear, but Pta1 and Pcf11 seem to be involved (Seoane et al. 2009).

Gene Regulation at the 3′ End

In addition to APA, mRNA 3′ processing can participate in gene regulation in other ways. The efficiency of mRNA 3′ processing plays an important role in controlling the mRNA abundance. mRNA 3′ processing factors can modulate other cellular processes to influence gene expression. In this section, we discuss a couple of such examples.

mRNA 3′ Processing in the Regulation of S. pombe Meiotic Genes

In *S. pombe*, the mRNAs for meiotic genes are not detectable in vegetative cells (Harigaya et al. 2006). However, several lines of evidence suggest that meiotic genes are transcribed, but are actively degraded. First, depletion of the exosome subunit Rrp6 results in the accumulation of hyperpolyadenylated meiotic mRNAs, and hyperpolyadenylation of these mRNAs in the Rrp6-depleted cells depends on Mmi1 (Meiotic mRNA interception 1), the poly(A) polymerase Pla1 (Harigaya et al. 2006; Yamanaka et al. 2010), and Red1, a CCCH zinc-finger-containing protein interacting with Mmi1 (Sugiyama and Sugioka-Sugiyama 2011). Second, depletion of the mRNA 3′ processing factors Rna15, Pla1 Pab2, Pfs2, and Dhp1, as well as several transcription termination factors all induce the accumulation of meiotic mRNAs (McPheeters et al. 2009; Yamanaka et al. 2010; St-Andre et al. 2010; Chen et al. 2011). Lastly, polyadenylation of meiotic mRNAs was shown to be required for their elimination in vegetative cells (McPheeters et al. 2009; Yamanaka et al. 2010).

Mmi1 plays a central role in this regulation (Fig. 3.5) (Harigaya et al. 2006). In mitotic cells, Mmi1 interacts with meiotic mRNAs with a specific cis-element containing the degenerate hexanucleotide motif UNAAAC (Zhang et al. 2010; Hiriart et al. 2012; Yamashita et al. 2012). Mmi1 recruits mRNA 3′ processing

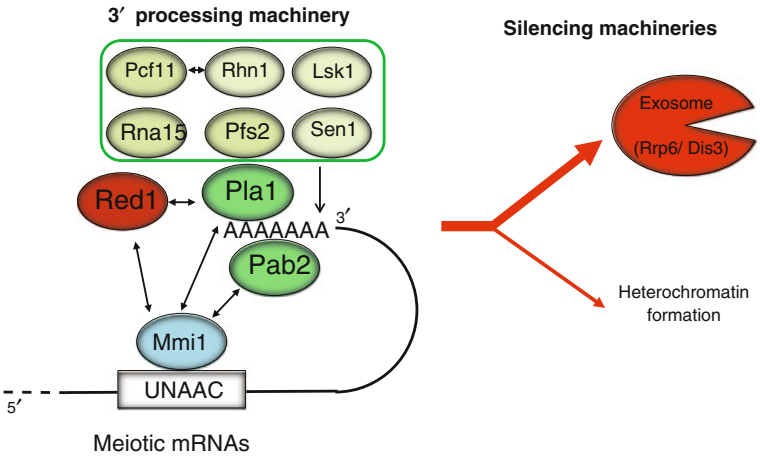


Fig. 3.5 mRNA 3′ processing factors are involved in the suppression of meiotic mRNAs. In vegetative cells, Mmi1 binds to UNAAC motifs in its target meiotic mRNAs and recruits mRNA 3′ processing factors, leading to polyadenylation of its target mRNAs and their degradation by the exosome. Mmi1-mediated meiotic mRNA degradation also promotes the formation of heterochromatin at meiotic gene loci

factors including Rna15 and Pla1 to hyperpolyadenylate its mRNA targets, which in turn are degraded by the exosome (Yamanaka et al. 2010). Additionally, Mmi-mediated meiotic mRNA degradation promotes the formation of heterochromatin at meiotic genes, which also contributes to meiotic gene silencing (Zofall et al. 2012). During meiosis, Mmi1 is sequestered by the master meiotic regulator Mei2 and the meiotic mRNAs are derepressed (Harigaya et al. 2006). Although mRNA 3′ processing factors are clearly involved in this regulatory pathway, their specific functions in this process remain unclear.

mRNA 3′ Processing Regulates the Expression of Neighboring Genes

Given the compressed nature of the yeast genomes, transcription read-through due to inefficient mRNA 3′ processing is likely to interfere with the expression of neighboring genes. When adjacent genes are arranged in tandem, transcription read-through from upstream genes may inhibit the transcription of downstream genes (Shearwin et al. 2005). On the other hand, when the neighboring genes are convergent, 3′ processing defects may lead to the collision of the transcription machinery (Prescott and Proudfoot 2002). Additionally, a recent study revealed an additional mechanism in *S. pombe* that regulate convergent gene expression in a cell cycle-dependent manner (Gullerova and Proudfoot 2008, 2012). During G1-S

phases, inefficient transcription termination leads to transcription read-through. The resulting double-stranded RNAs formed between the transcripts of convergent genes lead to the activation of the RNAi pathway. A transient RNAi-dependent heterochromatin structure is formed in the intergenic region between the convergent genes, characterized by the histone modification H3K9me3 and Swi6 binding, both hallmarks of heterochromatin (for more details see Chap. 13). Through a direct interaction, Swi6 induces the recruitment of cohesins at the chromatin of these convergent genes in the G2 phase. Cohesins are proteins involved in the regulation of sister chromatid separation during cell division. The presence of cohesins between convergent genes blocks transcription read-through and promotes transcription termination at the proper PAS, thereby restoring the mRNA levels of the convergent genes. After mitosis, cohesins are released and the heterochromatin structure at these loci is relieved (Gullerova and Proudfoot 2008, 2012). In *S. cerevisiae*, cohesins were also shown to concentrate at intergenic regions of convergent genes (Lengronne et al. 2004), but as H3K9me3 and RNAi are not conserved in budding yeast, this process might involve a different mechanism. This cell cycle-dependent gene regulation involving transcription termination regulation has been shown to be particularly important for the regulation of RNAi genes as 80 % of RNAi genes are convergent. This process of autoregulation may be important for the regulation of heterochromatin formation during different phases of the cell cycle (Gullerova and Proudfoot 2012; Zofall et al. 2012).

Conclusion /Future Directions

Studies in fungi have made tremendous contribution to our understanding of eukaryotic mRNA 3′ processing. Given the genetic tractability and the advent of high throughput analysis approaches, fungi will prove highly useful in addressing the remaining important questions in the field. First, although the list of essential mRNA 3′ processing factors is nearly complete, the functions of each factor in 3′ processing remains poorly characterized. The combination of genetic and biochemical analyses will be key to address this question. Second, as mentioned earlier, the mRNA 3′ processing machinery has evolved quite significantly in eukaryotes. Both PAS sequences and mRNA 3′ processing factors have diverged in different lineages and species. Given the increasingly number of species with their genomes sequenced, fungi provide a unique system to study the evolution of the mRNA 3′ processing system. Third, APA has increasingly been recognized as an important mechanism for gene regulation. Since APA is widespread in yeast and some of the regulatory mechanisms seem highly conserved between yeast and mammals, fungi again will be very useful in deciphering the "polyadenylation code," the rules by which PAS selection is regulated.

References

Amrani N, Minet M, Le Gouar M, Lacroute F, Wyers F (1997) Yeast Pab1 interacts with Rna15 and participates in the control of the poly(A) tail length in vitro. Mol Cell Biol 17(7): 3694–3701

Anderson JT, Wilson SM, Datar KV, Swanson MS (1993) NAB2: a yeast nuclear polyadenylated RNA-binding protein essential for cell viability. Mol Cell Biol 13(5):2730–2741

Aranda A, Proudfoot N (2001) Transcriptional termination factors for RNA polymerase II in yeast. Mol Cell 7(5):1003–1011

Barabino SM, Hubner W, Jenny A, Minvielle-Sebastia L, Keller W (1997) The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins. Genes Dev 11(13):1703–1716

Barilla D, Lee BA, Proudfoot NJ (2001) Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in Saccharomyces cerevisiae. Proc Natl Acad Sci USA 98(2):445–450

Beggs S, James TC, Bond U (2012) The PolyA tail length of yeast histone mRNAs varies during the cell cycle and is influenced by Sen1p and Rrp6p. Nucleic Acids Res 40(6):2700–2711

Bentley DL (2005) Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. Curr Opin Cell Biol 17(3):251–256

Birse CE, Lee BA, Hansen K, Proudfoot NJ (1997) Transcriptional termination signals for RNA polymerase II in fission yeast. EMBO J 16(12):3633–3643

Bougie I, Bisaillon M (2007) Characterization of the RNA binding energetics of the Candida albicans poly(A) polymerase. Yeast 24(5):431–446

Bucheli ME, Buratowski S (2005) Npl3 is an antagonist of mRNA 3' end formation by RNA polymerase II. EMBO J 24(12):2150–2160

Buratowski S (2003) The CTD code. Nat Struct Biol 10(9):679–680

Canavan R, Bond U (2007) Deletion of the nuclear exosome component RRP6 leads to continued accumulation of the histone mRNA HTB1 in S-phase of the cell cycle in *Saccharomyces cerevisiae*. Nucleic Acids Res 35(18):6268–6279

Chakraborty S, Sarmah B, Chakraborty N, Datta A (2002) Premature termination of RNA polymerase II mediated transcription of a seed protein gene in *Schizosaccharomyces pombe*. Nucleic Acids Res 30(13):2940–2949

Chan S, Choi EA, Shi Y (2011) Pre-mRNA 3'-end processing complex assembly and function. Wiley Interdiscip Rev RNA 2(3):321–335

Chen HM, Futcher B, Leatherwood J (2011) The fission yeast RNA binding protein Mmi1 regulates meiotic genes by controlling intron specific splicing and polyadenylation coupled RNA turnover. PLoS ONE 6(10):e26804

Chen S, Hyman LE (1998) A specific RNA–protein interaction at yeast polyadenylation efficiency elements. Nucleic Acids Res 26(21):4965–4974

Cheng H, He X, Moore C (2004) The essential WD repeat protein Swd2 has dual functions in RNA polymerase II transcription termination and lysine 4 methylation of histone H3. Mol Cell Biol 24(7):2932–2943

Colgan DF, Manley JL (1997) Mechanism and regulation of mRNA polyadenylation. Genes Dev 11(21):2755–2766

Connelly S, Manley JL (1988) A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. Genes Dev 2(4):440–452

Cremona N, Potter K, Wise JA (2011) A meiotic gene regulatory cascade driven by alternative fates for newly synthesized transcripts. Mol Biol Cell 22(1):66–77

Danckwardt S, Hentze MW, Kulozik AE (2008) 3' end mRNA processing: molecular mechanisms and implications for health and disease. EMBO J 27(3):482–498

Darmon SK, Lutz CS (2012) mRNA 3' end processing factors: a phylogenetic comparison. Comp Funct Genomics 2012:876893

de Klerk E, Venema A, Anvar SY, Goeman JJ, Hu O, den Dunnen JT, van der Maarel SM, Raz V, t Hoen PA (2012) Poly(A) binding protein nuclear 1 levels affect alternative polyadenylation. Nucleic Acids Res 40(18):9089–9101

Deka P, Bucheli ME, Moore C, Buratowski S, Varani G (2008) Structure of the yeast SR protein Npl3 and Interaction with mRNA 3'-end processing signals. J Mol Biol 375(1):136–150

Dermody JL, Dreyfuss JM, Villen J, Ogundipe B, Gygi SP, Park PJ, Ponticelli AS, Moore CL, Buratowski S, Bucheli ME (2008) Unphosphorylated SR-like protein Npl3 stimulates RNA polymerase II elongation. PLoS ONE 3(9):e3273

Derti A, Garrett-Engele P, Macisaac KD, Stevens RC, Sriram S, Chen R, Rohl CA, Johnson JM, Babak T (2012) A quantitative atlas of polyadenylation in five mammals. Genome Res 22(6):1173–1183

Di Giammartino DC, Nishida K, Manley JL (2011) Mechanisms and consequences of alternative polyadenylation. Mol Cell 43(6):853–866

Dichtl B, Blank D, Ohnacker M, Friedlein A, Roeder D, Langen H, Keller W (2002a) A role for SSU72 in balancing RNA polymerase II transcription elongation and termination. Mol Cell 10(5):1139–1150

Dichtl B, Blank D, Sadowski M, Hubner W, Weiser S, Keller W (2002b) Yhh1p/Cft1p directly links poly(A) site recognition and RNA polymerase II transcription termination. EMBO J 21(15):4125–4135

Dichtl B, Keller W (2001) Recognition of polyadenylation sites in yeast pre-mRNAs by cleavage and polyadenylation factor. EMBO J 20(12):3197–3209

Eissenberg JC, Shilatifard A (2010) Histone H3 lysine 4 (H3K4) methylation in development and differentiation. Dev Biol 339(2):240–249

Elkon R, Ugalde AP, Agami R (2013) Alternative cleavage and polyadenylation: extent, regulation and function. Nat Rev Genet 14(7):496–506

Ezeokonkwo C, Ghazy MA, Zhelkovsky A, Yeh PC, Moore C (2012) Novel interactions at the essential N-terminus of poly(A) polymerase that could regulate poly(A) addition in *Saccharomyces cerevisiae*. FEBS Lett 586(8):1173–1178

Fahrner K, Yarger J, Hereford L (1980) Yeast histone mRNA is polyadenylated. Nucleic Acids Res 8(23):5725–5737

Flavell SW, Kim TK, Gray JM, Harmin DA, Hemberg M, Hong EJ, Markenscoff-Papadimitriou E, Bear DM, Greenberg ME (2008) Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. Neuron 60(6):1022–1038

Franceschetti M, Bueno E, Wilson RA, Tucker SL, Gomez-Mena C, Calder G, Sesma A (2011) Fungal virulence and development is regulated by alternative pre-mRNA 3'-end processing in *Magnaporthe oryzae*. PLoS Pathog 7(12):e1002441

Freire-Picos MA, Hollenberg CP, Breunig KD, Cerdan ME (1995) Regulation of cytochrome c expression in the aerobic respiratory yeast *Kluyveromyces lactis*. FEBS Lett 360(1):39–42

Freire-Picos MA, Lombardia-Ferreira LJ, Ramil E, Gonzalez-Dominguez M, Cerdan ME (2001) The KlCYC1 gene, a downstream region for two differentially regulated transcripts. Yeast 18(14):1347–1355

Graber JH, Nazeer FI, Yeh PC, Kuehner JN, Borikar S, Hoskinson D, Moore CL (2013) DNA damage induces targeted, genome-wide variation of poly(A) sites in budding yeast. Genome Res 23(10):1690–1703

Graveley BR (2000) Sorting out the complexity of SR protein functions. RNA 6(9):1197–1211

Green DM, Johnson CP, Hagan H, Corbett AH (2003) The C-terminal domain of myosin-like protein 1 (Mlp1p) is a docking site for heterogeneous nuclear ribonucleoproteins that are required for mRNA export. Proc Natl Acad Sci USA 100(3):1010–1015

Gross S, Moore CL (2001) Rna15 interaction with the A-rich yeast polyadenylation signal is an essential step in mRNA 3'-end formation. Mol Cell Biol 21(23):8045–8055

Gullerova M, Proudfoot NJ (2008) Cohesin complex promotes transcriptional termination between convergent genes in *S. pombe*. Cell 132(6):983–995

Gullerova M, Proudfoot NJ (2012) Convergent transcription induces transcriptional gene silencing in fission yeast and mammalian cells. Nat Struct Mol Biol 19(11):1193–1201

Hansen K, Birse CE, Proudfoot NJ (1998) Nascent transcription from the nmt1 and nmt2 genes of *Schizosaccharomyces pombe* overlaps neighbouring genes. EMBO J 17(11):3066–3077

Harigaya Y, Tanaka H, Yamanaka S, Tanaka K, Watanabe Y, Tsutsumi C, Chikashige Y, Hiraoka Y, Yamashita A, Yamamoto M (2006) Selective elimination of messenger RNA prevents an incidence of untimely meiosis. Nature 442(7098):45–50

Hausmann S, Koiwa H, Krishnamurthy S, Hampsey M, Shuman S (2005) Different strategies for carboxyl-terminal domain (CTD) recognition by serine 5-specific CTD phosphatases. J Biol Chem 280(45):37681–37688

He X, Khan AU, Cheng H, Pappas DL Jr, Hampsey M, Moore CL (2003) Functional interactions between the transcription and mRNA 3' end processing machineries mediated by Ssu72 and Sub1. Genes Dev 17(8):1030–1042

He X, Moore C (2005) Regulation of yeast mRNA 3' end processing by phosphorylation. Mol Cell 19(5):619–629

Hector RE, Nykamp KR, Dheur S, Anderson JT, Non PJ, Urbinati CR, Wilson SM, Minvielle-Sebastia L, Swanson MS (2002) Dual requirement for yeast hnRNP Nab2p in mRNA poly(A) tail length control and nuclear export. EMBO J 21(7):1800–1810

Helmling S, Zhelkovsky A, Moore CL (2001) Fip1 regulates the activity of Poly(A) polymerase through multiple interactions. Mol Cell Biol 21(6):2026–2037

Hiriart E, Vavasseur A, Touat-Todeschini L, Yamashita A, Gilquin B, Lambert E, Perot J, Shichino Y, Nazaret N, Boyault C, Lachuer J, Perazza D, Yamamoto M, Verdel A (2012) Mmi1 RNA surveillance machinery directs RNAi complex RITS to specific meiotic genes in fission yeast. EMBO J 31(10):2296–2308

Hirose Y, Manley JL (2000) RNA polymerase II and the integration of nuclear events. Genes Dev 14(12):1415–1429

Hoopes BC, Bowers GD, DiVisconte MJ (2000) The two Saccharomyces cerevisiae SUA7 (TFIIB) transcripts differ at the 3'-end and respond differently to stress. Nucleic Acids Res 28(22):4435–4443

Humphrey T, Birse CE, Proudfoot NJ (1994) RNA 3' end signals of the *S. pombe* ura4 gene comprise a site determining and efficiency element. EMBO J 13(10):2441–2451

Humphrey T, Christofori G, Lucijanic V, Keller W (1987) Cleavage and polyadenylation of messenger RNA precursors in vitro occurs within large and specific 3' processing complexes. EMBO J 6(13):4159–4168

Humphrey T, Sadhale P, Platt T, Proudfoot N (1991) Homologous mRNA 3' end formation in fission and budding yeast. EMBO J 10(11):3503–3511

Jenal M, Elkon R, Loayza-Puch F, van Haaften G, Kuhn U, Menzies FM, Oude Vrielink JA, Bos AJ, Drost J, Rooijers K, Rubinsztein DC, Agami R (2012) The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. Cell 149(3):538–553

Johnson SA, Cubberley G, Bentley DL (2009) Cotranscriptional recruitment of the mRNA export factor Yra1 by direct interaction with the 3' end processing factor Pcf11. Mol Cell 33(2):215–226

Johnson SA, Kim H, Erickson B, Bentley DL (2011) The export factor Yra1 modulates mRNA 3' end processing. Nat Struct Mol Biol 18(10):1164–1171

Kerr SC, Azzouz N, Fuchs SM, Collart MA, Strahl BD, Corbett AH, Laribee RN (2013) The Ccr4-Not complex interacts with the mRNA export machinery. PLoS ONE 6(3):e18302

Kessler MM, Henry MF, Shen E, Zhao J, Gross S, Silver PA, Moore CL (1997) Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3'-end formation in yeast. Genes Dev 11(19):2545–2556

Kessler MM, Zhao J, Moore CL (1996) Purification of the *Saccharomyces cerevisiae* cleavage/polyadenylation factor I. Separation into two components that are required for both cleavage and polyadenylation of mRNA 3' ends. J Biol Chem 271(43):27167–27175

Kim Guisbert KS, Li H, Guthrie C (2007) Alternative 3' pre-mRNA processing in *Saccharomyces cerevisiae* is modulated by Nab4/Hrp1 in vivo. PLoS Biol 5(1):e6

Kim M, Krogan NJ, Vasiljeva L, Rando OJ, Nedea E, Greenblatt JF, Buratowski S (2004) The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature 432(7016):517–522

Kleiman FE, Manley JL (2001) The BARD1-CstF-50 interaction links mRNA 3' end formation to DNA damage and tumor suppression. Cell 104(5):743–753

Kress TL, Krogan NJ, Guthrie C (2008) A single SR-like protein, Npl3, promotes pre-mRNA splicing in budding yeast. Mol Cell 32(5):727–734

Krishnamurthy S, He X, Reyes-Reyes M, Moore C, Hampsey M (2004) Ssu72 is an RNA polymerase II CTD phosphatase. Mol Cell 14(3):387–394

Kyburz A (2003) The role of the yeast cleavage and polyadenylation factor subunit Ydh1p/Cft2p in pre-mRNA 3'-end formation. Nucleic Acids Res 31(14):3936–3945

Lee JY, Yeh I, Park JY, Tian B (2007) PolyA_DB 2: mRNA polyadenylation sites in vertebrate genes. Nucleic Acids Res 35(Database issue):D165–D168.

Leeper TC, Qu X, Lu C, Moore C, Varani G (2010) Novel protein–protein contacts facilitate mRNA 3'-processing signal recognition by Rna15 and Hrp1. J Mol Biol 401(3):334–349

Lengronne A, Katou Y, Mori S, Yokobayashi S, Kelly GP, Itoh T, Watanabe Y, Shirahige K, Uhlmann F (2004) Cohesin relocation from sites of chromosomal loading to places of convergent transcription. Nature 430(6999):573–578

Licatalosi DD, Geiger G, Minet M, Schroeder S, Cilli K, McNeil JB, Bentley DL (2002) Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. Mol Cell 9(5):1101–1111

Lingner J, Radtke I, Wahle E, Keller W (1991) Purification and characterization of poly(A) polymerase from *Saccharomyces cerevisiae*. J Biol Chem 266(14):8741–8746

Lund MK, Kress TL, Guthrie C (2008) Autoregulation of Npl3, a yeast SR protein, requires a novel downstream region and serine phosphorylation. Mol Cell Biol 28(11):3873–3881

MacDonald CC, Wilusz J, Shenk T (1994) The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. Mol Cell Biol 14(10):6647–6654

Mandart E, Parker R (1995) Effects of mutations in the *Saccharomyces cerevisiae* RNA14, RNA15, and PAP1 genes on polyadenylation in vivo. Mol Cell Biol 15(12):6979–6986

Mandel CR, Bai Y, Tong L (2008) Protein factors in pre-mRNA 3'-end processing. Cell Mol Life Sci 65(7–8):1099–1122

Mandel CR, Kaneko S, Zhang H, Gebauer D, Vethantham V, Manley JL, Tong L (2006) Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. Nature 444(7121):953–956

Mangus DA, Evans MC, Agrin NS, Smith M, Gongidi P, Jacobson A (2004) Positive and negative regulation of poly(A) nuclease. Mol Cell Biol 24(12):5521–5533

Mangus DA, Evans MC, Jacobson A (2003) Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. Genome Biol 4(7):223

Marfatia KA, Crafton EB, Green DM, Corbett AH (2003) Domain analysis of the *Saccharomyces cerevisiae* heterogeneous nuclear ribonucleoprotein, Nab2p. Dissecting the requirements for Nab2p-facilitated poly(A) RNA export. J Biol Chem 278(9):6731–6740

Marzluff WF, Wagner EJ, Duronio RJ (2008) Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nat Rev Genet 9(11):843–854

Mata J (2013) Genome-wide mapping of polyadenylation sites in fission yeast reveals widespread alternative polyadenylation. RNA Biol 10(8):1407–1414

Mayer SA, Dieckmann CL (1989) The yeast CBP1 gene produces two differentially regulated transcripts by alternative 3'-end formation. Mol Cell Biol 9(10):4161–4169

Mayer SA, Dieckmann CL (1991) Yeast CBP1 mRNA 3' end formation is regulated during the induction of mitochondrial function. Mol Cell Biol 11(2):813–821

Mayr C, Bartel DP (2009) Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell 138(4):673–684

McPheeters DS, Cremona N, Sunder S, Chen HM, Averbeck N, Leatherwood J, Wise JA (2009) A complex gene regulatory mechanism that operates at the nexus of multiple RNA processing decisions. Nat Struct Mol Biol16(3):255–264

Meinhart A, Cramer P (2004) Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. Nature 430(6996):223–226

Meinke G, Ezeokonkwo C, Balbo P, Stafford W, Moore C, Bohm A (2008) Structure of yeast poly(A) polymerase in complex with a peptide from Fip1, an intrinsically disordered protein. Biochemistry 47(26):6859–6869

Miller T, Krogan NJ, Dover J, Erdjument-Bromage H, Tempst P, Johnston M, Greenblatt JF, Shilatifard A (2001) COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. Proc Natl Acad Sci USA 98(23):12902–12907

Minvielle-Sebastia L, Preker PJ, Keller W (1994) RNA14 and RNA15 proteins as components of a yeast pre-mRNA 3'-end processing factor. Science 266(5191):1702–1705

Minvielle-Sebastia L, Preker PJ, Wiederkehr T, Strahm Y, Keller W (1997) The major yeast poly(A)-binding protein is associated with cleavage factor IA and functions in premessenger RNA 3'-end formation. Proc Natl Acad Sci USA 94(15):7897–7902

Mizrahi N, Moore C (2000) Posttranslational phosphorylation and ubiquitination of the *Saccharomyces cerevisiae* poly(A) polymerase at the S/G2 Stage of the cell cycle. Mol Cell Biol 20(8):2794–2802

Moqtaderi Z, Geisberg JV, Jin Y, Fan X, Struhl K (2013) Species-specific factors mediate extensive heterogeneity of mRNA 3' ends in yeasts. Proc Natl Acad Sci USA 110(27): 11073–11078

Munsterkotter M, Steinberg G (2007) The fungus Ustilago maydis and humans share disease-related proteins that are not found in *Saccharomyces cerevisiae*. BMC Genom 8:473

Nedea E, He X, Kim M, Pootoolal J, Zhong G, Canadien V, Hughes T, Buratowski S, Moore CL, Greenblatt J (2003) Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3'-ends. J Biol Chem 278(35):33000–33010

Nedea E, Nalbant D, Xia D, Theoharis NT, Suter B, Richardson CJ, Tatchell K, Kislinger T, Greenblatt JF, Nagy PL (2008) The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes. Mol Cell 29(5):577–587

Noble CG, Beuth B, Taylor IA (2007) Structure of a nucleotide-bound Clp1-Pcf11 polyadenylation factor. Nucleic Acids Res 35(1):87–99

Noble CG, Walker PA, Calder LJ, Taylor IA (2004) Rna14–Rna15 assembly mediates the RNA-binding capability of *Saccharomyces cerevisiae* cleavage factor IA. Nucleic Acids Res 32(11):3364–3375

Ohnacker M, Barabino SM, Preker PJ, Keller W (2000) The WD-repeat protein pfs2p bridges two essential factors within the yeast pre-mRNA 3'-end-processing complex. EMBO J 19(1):37–47

Ohnacker M, Minvielle-Sebastia L, Keller W (1996) The *Schizosaccharomyces pombe* pla1 gene encodes a poly(A) polymerase and can functionally replace its *Saccharomyces cerevisiae* homologue. Nucleic Acids Res 24(13):2585–2591

Ozsolak F, Kapranov P, Foissac S, Kim SW, Fishilevich E, Monaghan AP, John B, Milos PM (2010) Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. Cell 143(6):1018–1029

Perez-Canadillas JM (2006) Grabbing the message: structural basis of mRNA 3'UTR recognition by Hrp1. EMBO J 25(13):3167–3178

Perreault A, Lemieux C, Bachand F (2007) Regulation of the nuclear poly(A)-binding protein by arginine methylation in fission yeast. J Biol Chem 282(10):7552–7562

Potter K, Cremona N, Sunder S, Wise JA (2012) A dominant role for meiosis-specific 3' RNA processing in controlling expression of a fission yeast cyclin gene. RNA 18(7):1408–1420

Preker PJ, Lingner J, Minvielle-Sebastia L, Keller W (1995) The FIP1 gene encodes a component of a yeast pre-mRNA polyadenylation factor that directly interacts with poly(A) polymerase. Cell 81(3):379–389

Prescott EM, Proudfoot NJ (2002) Transcriptional collision between convergent genes in budding yeast. Proc Natl Acad Sci USA 99(13):8796–8801

Proudfoot N (2004) New perspectives on connecting messenger RNA 3' end formation to transcription. Curr Opin Cell Biol 16(3):272–278

Proudfoot NJ (2011) Ending the message: poly(A) signals then and now. Genes Dev 25(17):1770–1782

Pugh DJ, Ab E, Faro A, Lutya PT, Hoffmann E, Rees DJ (2006) DWNN, a novel ubiquitin-like domain, implicates RBBP6 in mRNA processing and ubiquitin-like pathways. BMC Struct Biol 6:1

Qu X, Lykke-Andersen S, Nasser T, Saguez C, Bertrand E, Jensen TH, Moore C (2009) Assembly of an export-competent mRNP is needed for efficient release of the 3'-end processing complex after polyadenylation. Mol Cell Biol 29(19):5327–5338

Qu X, Perez-Canadillas JM, Agrawal S, De Baecke J, Cheng H, Varani G, Moore C (2007) The C-terminal domains of vertebrate CstF-64 and its yeast orthologue Rna15 form a new structure critical for mRNA 3'-end processing. J Biol Chem 282(3):2101–2115

Ramirez A, Shuman S, Schwer B (2008) Human RNA 5'-kinase (hClp1) can function as a tRNA splicing enzyme in vivo. RNA 14(9):1737–1745

Richter JD (1999) Cytoplasmic polyadenylation in development and beyond. Microbiol Mol Biol Rev 63(2):446–456

Rissland OS, Mikulasova A, Norbury CJ (2007) Efficient RNA polyuridylation by noncanonical poly(A) polymerases. Mol Cell Biol 27(10):3612–3624

Rodriguez-Torres AM, Lamas-Maceiras M, Garcia-Diaz R, Freire-Picos MA (2013) Structurally conserved and functionally divergent yeast Ssu72 phosphatases. FEBS Lett 587(16): 2617–2622

Roguev A, Schaft D, Shevchenko A, Pijnappel WW, Wilm M, Aasland R, Stewart AF (2001) The *Saccharomyces cerevisiae* Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. EMBO J 20(24):7137–7148

Russnak R, Nehrke KW, Platt T (1995) REF2 encodes an RNA-binding protein directly involved in yeast mRNA 3'-end formation. Mol Cell Biol 15(3):1689–1697

Ryan K, Calvo O, Manley JL (2004) Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. RNA 10(4):565–573

Sachs AB, Davis RW (1989) The poly(A) binding protein is required for poly(A) shortening and 60S ribosomal subunit-dependent translation initiation. Cell 58(5):857–867

Sakai Y, Saijo M, Coelho K, Kishino T, Niikawa N, Taya Y (1995) cDNA sequence and chromosomal localization of a novel human protein, RBQ-1 (RBBP6), that binds to the retinoblastoma gene product. Genomics 30(1):98–101

Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB (2008) Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. Science 320(5883):1643–1647

Schlackow M, Marguerat S, Proudfoot NJ, Bahler J, Erban R, Gullerova M (2013) Genome-wide analysis of poly(A) site selection in *Schizosaccharomyces pombe*. RNA 19(12):1617–1631

Schmid M, Poulsen MB, Olszewski P, Pelechano V, Saguez C, Gupta I, Steinmetz LM, Moore C, Jensen TH (2012) Rrp6p controls mRNA poly(A) tail length and its decoration with poly(A) binding proteins. Mol Cell 47(2):267–280

Scott DD, Norbury CJ (2013) RNA decay via 3' uridylation. Biochim Biophys Acta 1829(6–7): 654–665

Seoane S, Guiard B, Rodriguez-Torres AM, Freire-Picos MA (2005) Effects of splitting alternative KlCYC1 3'-UTR regions on processing: metabolic consequences and biotechnological applications. J Biotechnol 118(2):149–156

Seoane S, Lamas-Maceiras M, Rodriguez-Torres AM, Freire-Picos MA (2009) Involvement of Pta1, Pcf11 and a KlCYC1 AU-rich element in alternative RNA 3'-end processing selection in yeast. FEBS Lett 583(17):2843–2848

Shearwin KE, Callen BP, Egan JB (2005) Transcriptional interference–a crash course. Trends Genet 21(6):339–345

Shepard PJ, Choi E, Lu J, Flanagan LA, Hertel KJ, Shi Y (2011) Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. RNA 17(4):761–772

Shi Y (2012) Alternative polyadenylation: new insights from global analyses. RNA 18(12):2105–2117

Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR 3rd, Frank J, Manley JL (2009) Molecular architecture of the human pre-mRNA 3' processing complex. Mol Cell 33(3):365–376

Shobuike T, Tatebayashi K, Tani T, Sugano S, Ikeda H (2001) The dhp1(+) gene, encoding a putative nuclear 5' → 3' exoribonuclease, is required for proper chromosome segregation in fission yeast. Nucleic Acids Res 29(6):1326–1333

Skolnik-David H, Moore CL, Sharp PA (1987) Electrophoretic separation of polyadenylation-specific complexes. Genes Dev 1(7):672–682

Soucek S, Corbett AH, Fasken MB (2012) The long and the short of it: the role of the zinc finger polyadenosine RNA binding protein, Nab2, in control of poly(A) tail length. Biochim Biophys Acta 1819(6):546–554

Sparks KA, Dieckmann CL (1998) Regulation of poly(A) site choice of several yeast mRNAs. Nucleic Acids Res 26(20):4676–4687

St-Andre O, Lemieux C, Perreault A, Lackner DH, Bahler J, Bachand F (2010) Negative regulation of meiotic gene expression by the nuclear poly(a)-binding protein in fission yeast. J Biol Chem 285(36):27859–27868

Stumpf G, Domdey H (1996) Dependence of yeast pre-mRNA 3'-end processing on CFT1: a sequence homolog of the mammalian AAUAAA binding factor. Science 274(5292): 1517–1520

Sugano S, Shobuike T, Takeda T, Sugino A, Ikeda H (1994) Molecular analysis of the dhp1+ gene of Schizosaccharomyces pombe: an essential gene that has homology to the DST2 and RAT1 genes of *Saccharomyces cerevisiae*. Mol Gen Genet 243(1):1–8

Sugiyama T, Sugioka-Sugiyama R (2011) Red1 promotes the elimination of meiosis-specific mRNAs in vegetatively growing fission yeast. EMBO J 30(6):1027–1039

Tacahashi Y, Helmling S, Moore CL (2003) Functional dissection of the zinc finger and flanking domains of the Yth1 cleavage/polyadenylation factor. Nucleic Acids Res 31(6):1744–1752

Takagaki Y, Manley JL (1997) RNA recognition by the human polyadenylation factor CstF. Mol Cell Biol 17(7):3907–3914

Takagaki Y, Manley JL (2000) Complex protein interactions within the human polyadenylation machinery identify a novel component. Mol Cell Biol 20(5):1515–1525

Thakurta AG, Ho Yoon J, Dhar R (2002) Schizosaccharomyces pombe spPABP, a homologue of Saccharomyces cerevisiae Pab1p, is a non-essential, shuttling protein that facilitates mRNA export. Yeast 19(9):803–810

Tian B, Manley JL (2013) Alternative cleavage and polyadenylation: the long and short of it. Trends Biochem Sci 38(6):312–320

Vo LT, Minet M, Schmitter JM, Lacroute F, Wyers F (2001) Mpe1, a zinc knuckle protein, is an essential component of yeast cleavage and polyadenylation factor required for the cleavage and polyadenylation of mRNA. Mol Cell Biol 21(24):8346–8356

Wang SW, Asakawa K, Win TZ, Toda T, Norbury CJ (2005) Inactivation of the pre-mRNA cleavage and polyadenylation factor Pfs2 in fission yeast causes lethal cell cycle defects. Mol Cell Biol 25(6):2288–2296

Weitzer S, Martinez J (2007) The human RNA kinase hClp1 is active on 3' transfer RNA exons and short interfering RNAs. Nature 447(7141):222–226

West S, Gromak N, Proudfoot NJ (2004) Human 5' → 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. Nature 432(7016):522–525

Yamanaka S, Yamashita A, Harigaya Y, Iwata R, Yamamoto M (2010) Importance of polyadenylation in the selective elimination of meiotic mRNAs in growing *S. pombe* cells. EMBO J 29(13):2173–2181

Yamashita A, Shichino Y, Tanaka H, Hiriart E, Touat-Todeschini L, Vavasseur A, Ding DQ, Hiraoka Y, Verdel A, Yamamoto M (2012) Hexanucleotide motifs mediate recruitment of the RNA elimination machinery to silent meiotic genes. Open Biol 2(3):120014

Yang Q, Coseno M, Gilmartin GM, Doublie S (2010) Crystal structure of a human cleavage factor CFI(m)25/CFI(m)68/RNA complex provides an insight into poly(A) site recognition and RNA looping. Structure 19(3):368–377

Yao C, Biesinger J, Wan J, Weng L, Xing Y, Xie X, Shi Y (2012) Transcriptome-wide analyses of CstF64-RNA interactions in global regulation of mRNA alternative polyadenylation. Proc Natl Acad Sci USA 109(46):18773–18778

Yoon OK, Brem RB (2010) Noncanonical transcript forms in yeast and their regulation during environmental stress. RNA 16(6):1256–1267

Yu H, Braun P, Yildirim MA, Lemmens I, Venkatesan K, Sahalie J, Hirozane-Kishikawa T, Gebreab F, Li N, Simonis N, Hao T, Rual JF, Dricot A, Vazquez A, Murray RR, Simon C, Tardivo L, Tam S, Svrzikapa N, Fan C, de Smet AS, Motyl A, Hudson ME, Park J, Xin X, Cusick ME, Moore T, Boone C, Snyder M, Roth FP, Barabasi AL, Tavernier J, Hill DE, Vidal M (2008) High-quality binary protein interaction map of the yeast interactome network. Science 322(5898):104–110

Yu L, Volkert MR (2013) UV damage regulates alternative polyadenylation of the RPB2 gene in yeast. Nucleic Acids Res 41(5):3104–3114

Zhang Z, Gilmour DS (2006) Pcf11 is a termination factor in Drosophila that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. Mol Cell 21(1):65–74

Zhang Z, Klatt A, Henderson AJ, Gilmour DS (2007) Transcription termination factor Pcf11 limits the processivity of Pol II on an HIV provirus to repress gene expression. Genes Dev 21(13):1609–1614

Zhang Z, Theler D, Kaminska KH, Hiller M, de la Grange P, Pudimat R, Rafalska I, Heinrich B, Bujnicki JM, Allain FH, Stamm S (2010) The YTH domain is a novel RNA binding domain. J Biol Chem 285(19):14701–14710

Zhao J, Hyman L, Moore C (1999) Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. Microbiol Mol Biol Rev 63(2):405–445

Zhao J, Kessler MM, Moore CL (1997) Cleavage factor II of Saccharomyces cerevisiae contains homologues to subunits of the mammalian Cleavage/ polyadenylation specificity factor and exhibits sequence-specific, ATP-dependent interaction with precursor RNA. J Biol Chem 272(16):10831–10838

Zhelkovsky A, Tacahashi Y, Nasser T, He X, Sterzer U, Jensen TH, Domdey H, Moore C (2006) The role of the Brr5/Ysh1 C-terminal domain and its homolog Syc1 in mRNA 3'-end processing in *Saccharomyces cerevisiae*. RNA 12(3):435–445

Zofall M, Yamanaka S, Reyes-Turcu FE, Zhang K, Rubin C, Grewal SI (2012) RNA elimination machinery targeting meiotic mRNAs promotes facultative heterochromatin formation. Science 335(6064):96–100

Chapter 4
mRNA Export

Alexandra Hackmann and Heike Krebber

Abstract Fungi, like all other eukaryotic cells, have a separated place of transcription and translation. Compartmentalization into nucleus and cytoplasm by the nuclear envelope necessitates bidirectional traffic of small molecules, proteins, and large RNA-protein macromolecules, which pass through nuclear pore complexes (NPCs). One major nucleo-cytoplasmic transport process is the mRNA export, not only because of the high amount of permanently generated transcripts, but also because of its enormous size. In contrast to the protein transport and the translocation of small RNAs, such as tRNAs or spliceosomal UsnRNAs, that involve the karyopherins as transport receptors and the Ran GTPase system as the driving force, bulk mRNA export requires other factors. Instead the highly conserved mRNA export receptor heterodimer *Mex67-Mtr2* (*NXF1-NXT1* or TAP-p15 in metazoans) is recruited to the mRNA and contacts the nucleoporins of the NPC to allow transit. Directionality of the transport event is provided by the ATP-dependent remodeling of the RNA/protein complexes by the DEAD-box RNA helicase Dbp5, which acts on the cytoplasmic side of the NPC. mRNA export is tightly coupled to transcription and mRNA maturation and the whole process is surveyed by a nuclear surveillance machinery that prevents immature and false transcripts from slippage into the cytoplasm and their consequent translation. These general processes of transcription, processing, and mRNA export are highly conserved among all eukaryotes, including all members of the fungal kingdom. However, the best-studied organism is the budding yeast *Saccharomyces cerevisiae* and therefore this book chapter will mostly focus on this organism.

A. Hackmann · H. Krebber (✉)
Abteilung für Molekulare Genetik, Institut für Mikrobiologie und Genetik,
Göttinger Zentrum für Molekulare Biowissenschaften (GZMB),
Georg-August Universität Göttingen, Göttingen, Germany
e-mail: heike.krebber@biologie.uni-goettingen.de

A. Sesma and T. von der Haar (eds.), *Fungal RNA Biology*,
DOI: 10.1007/978-3-319-05687-6_4,

Contents

The Early Phase: From Pre-mRNA Transcription and Maturation to Formation of an Export Competent mRNP

All sequential steps in pre-mRNA transcription and processing are interconnected and finally lead to the recruitment of factors that are necessary for proper packaging and export of the mature messenger ribonucleoparticle (mRNP) (Fig. 4.1; see also Chap. 1). During transcript synthesis by RNA polymerase II (RNA pol II), different modifications of its C-terminal domain (CTD) lead to the timely recruitment of mRNA processing factors, such as 5′ capping, splicing, 3′-end cleavage, polyadenylation and export factors, which themselves are influenced in their functions by different modifications (Bentley 2005; Tutucci and Stutz 2011; Hsin and Manley 2012; Heidemann et al. 2013).

Initially, RNA pol II in its unphosphorylated status associates with the promoter region and forms the pre-initiation complex (Buratowski 2009; Bataille et al. 2012; Heidemann et al. 2013). Upon synthesis of the first few nucleotides of the pre-mRNA, serine five phosphorylation of the CTD leads to the recruitment of 5′ RNA capping enzymes that start pre-mRNA processing (Lidschreiber et al. 2013). As a consequence, the 5′-triphosphate end of the pre-mRNA receives a protective 7-methylguanosine cap, which is subsequently recognized and bound by *Cbp80* and *Cbp20* that form the cap binding complex (CBC). This cap structure not only protects the mRNA from degradation but also influences spliceosome assembly and impedes premature transcription termination of cryptic termination sites by preventing the loading of 3′-end processing factors (Colot et al. 1996; Gornemann et al. 2005; Wong et al. 2007).

The shuttling serine/arginine (SR)-rich protein *Npl3* is the first one in the line of several mRNA binding proteins that assemble on the nascent transcript and promote the export of the mature transcript. *Npl3* is transferred to the pre-mRNA upon serine 2 phosphorylation of the CTD of RNA pol II during transcription initiation (Lei et al. 2001; Dermody et al. 2008). Both the phosphorylation status of the CTD and the presence of *Npl3* on the pre-mRNA have an impact on early splicing factor assembly

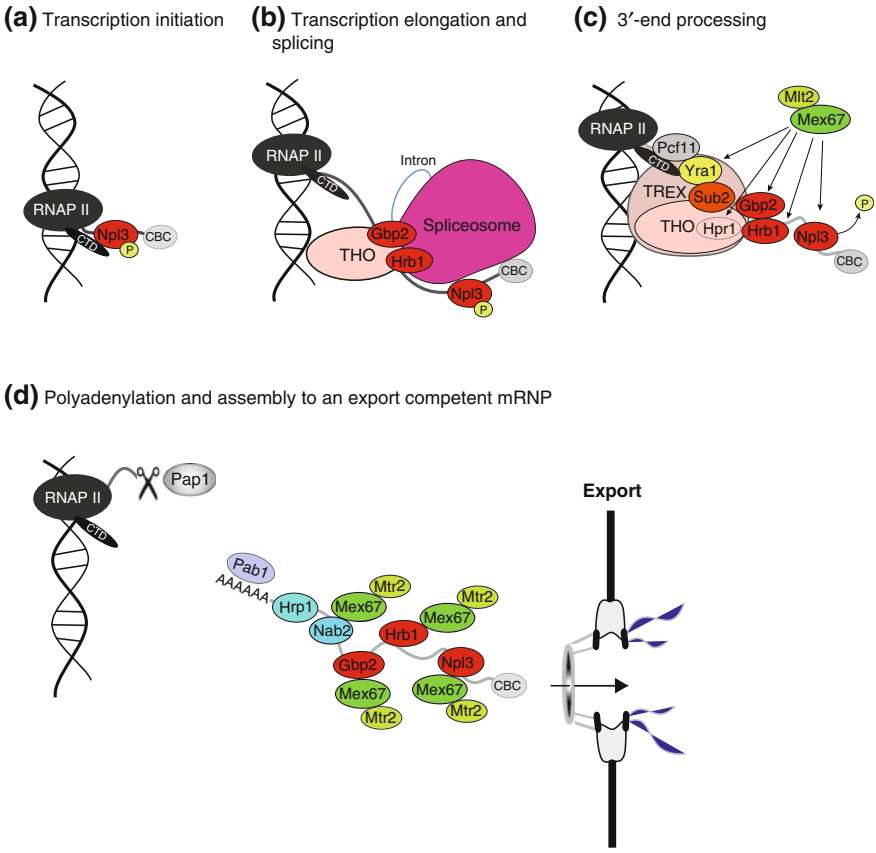


Fig. 4.1 Co-transcriptional processing and assembly of an export competent mRNP. **a** RNA polymerase II (RNAP II) promotes association of mRNA binding proteins such as the CBC and Npl3 during transcription initiation. **b** The THO complex promotes transcription elongation. Splicing occurs co-transcriptionally. *Gbp2* and *Hrb1* bind to the mRNP during splicing. **c** The THO complex interacts with *Sub2* and *Yra1* to form the TREX complex and subsequent 3′-end processing occurs. **d** Polyadenylation and recruitment of *Nab2*, *Hrp1* and *Pab1* takes place. Upon completion of the loading of the export receptor heterodimer *Mex67-Mtr2* onto the mRNP it becomes export competent and can leave the nucleus

and thus proper splicing (Morris and Greenleaf 2000; Kress et al. 2008). On the matured mRNA *Npl3* is one of several proteins that bind the essential mRNA export receptor heterodimer *Mex67-Mtr2* that is required for the passage through the hydrophobic meshwork of the NPC (Lee et al. 1996; Strasser et al. 2000; Gilbert and Guthrie 2004). Consequently, mutations in *NPL3* and *MEX67* or *MTR2* lead to mRNA export defects (Kadowaki et al. 1994; Lee et al. 1996; Hurt et al. 2000).

During transcription elongation, the phosphorylation status of the CTD changes again and RNA pol II recruits another important complex: the heterotetrameric THO complex. THO supports RNA pol II progress by preventing DNA:RNA hybrid formation during transcription elongation (Huertas and Aguilera 2003;

Rondon et al. 2003). Mutations in components of the THO complex do not only impact transcription elongation but also the export of mature mRNAs as it is involved in the recruitment of several proteins to the mRNA (Strasser et al. 2002; Hurt et al. 2004; Gwizdek et al. 2006).

Sub2 and *Yra1* bind to the THO complex on the nascent mRNA to form the *TR*anscription-*EX*port (TREX)-complex (Strasser et al. 2002; Rodriguez-Navarro and Hurt 2011). *Ch*romatin *i*mmuno*p*recipitation experiments (ChIPs) indicated that *Yra1* loading occurs during transcription elongation (Lei et al. 2001; Johnson et al. 2009). In fact, TREX complex formation depends on hyperphosphorylation of the CTD of RNA pol II, as shown by interaction studies with *Yra1* (MacKellar and Greenleaf 2011). Later, *Yra1* and the ubiquitinated THO/TREX component *Hpr1* recruit *Mex67* for proper export (Strasser and Hurt 2000; Gwizdek et al. 2006; Babour et al. 2012; Katahira 2012). Consequently, mutations in the TREX complex components lead to mRNA export defects (Jensen et al. 2001a; Rodriguez-Navarro et al. 2002; Strasser et al. 2002). Interestingly, the Cap structure might support efficient nuclear export of the mature mRNA by proper positioning of the TREX complex (Lewis and Izaurralde 1997; Katahira 2012).

Splicing occurs on intron containing mRNAs co-transcriptionally (see Chap. 2). For more details on the splicing process, some excellent recent reviews are suggested (Meyer and Vilardell 2009; Wahl et al. 2009; Will and Luhrmann 2011; Chen and Cheng 2012). Besides its well-established function in splicing the *Prp19*-complex, which has been identified as a non-snRNP constituent of the spliceosome, was suggested to function additionally in general transcription elongation, because *Prp19* is loaded by the elongating RNA pol II onto intron-containing and intronless transcripts (Tarn et al. 1994; Chanarat et al. 2011). Its proper binding has been shown to be a prerequisite for the TREX complex recruitment (Chanarat et al. 2011, 2012). This reflects the strong interplay between transcription, splicing, and export.

During splicing, two other shuttling SR-proteins are loaded onto the pre-mRNA, termed *Gbp2* and *Hrb1* (Windgassen and Krebber 2003; Hacker and Krebber 2004; Hackmann et al. 2014). In contrast to *Npl3*, *Gbp2* and *Hrb1* are recruited in a THO complex and splicing-dependent manner and interact with *Mex67* (Hacker and Krebber 2004; Hurt et al. 2004; Hackmann et al. 2014).

After capping and splicing, the transcript is finally cleaved and the 3′-end is polyadenylated (Barilla et al. 2001; Ahn et al. 2004). Recruitment of 3′-end processing and polyadenylation factors was shown to require the Ctk1 mediated serine 2 phosphorylation of the RNA pol II CTD (Ahn et al. 2004). Transcription of the polyadenylation signal sequence in the 3′UTR initiates this final pre-mRNA processing step and is recognized by the 3′-end processing machinery (see Chap. 3). This AU-rich sequence is further recognized by *Hrp1*, which is loaded to support the efficiency of the cleavage reaction (Perez-Canadillas 2006; Barnwal et al. 2012). Upon 3′- end processing, *Hrp1* remains bound to the RNA and shuttles with the mature mRNA to the cytoplasm, where it is further involved in the nonsense mediated decay (NMD) of premature stop-codon containing mRNAs (Gonzáles et al. 2000).

Upon cleavage the poly(A) polymerase Pap1 (*Pla1* in *Schizosaccharomyces pombe*) conducts polyadenylation of the upstream cleavage product. Recruitment of the poly(A) binding proteins *Nab2* and *Pab1* leads to a controlled poly(A) tail length (Anderson et al. 1993; Hector et al. 2002; Dunn et al. 2005). This final processing step triggers the displacement of several associated processing factors such as the cleavage and polyadenylation factors and TREX, which leads to the release of the mRNP from the transcription apparatus (Kim et al. 2004; Dunn et al. 2005; Qu et al. 2009). One more time *Mex67-Mtr2* associates with the mRNP by interaction with *Nab2* and the mRNP is now ready for being exported (Green et al. 2002; Batisse et al. 2009). The mature mRNP is now covered with several *Mex67-Mtr2* molecules, which are loaded at different positions, through interaction with *Npl3*, *Gbp2*, *Hrb1*, *Hpr1*, *Yra1*, and *Nab2* (Strasser and Hurt 2000; Green et al. 2002; Gwizdek et al. 2006; Hackmann et al. 2014). These mRNA export adapters for *Mex67-Mtr2* are recruited at individual processing steps and it seems likely that every maturation step is controlled and finally flagged with one or more *Mex67-Mtr2* molecules (Hackmann et al. 2014). Moreover, post-translational modifications such as ubiquitination of *Hpr1* or dephosphorylation of *Npl3* also have an impact on the association of *Mex67-Mtr2* (Strasser and Hurt 2000; Gilbert and Guthrie 2004). This might prevent a too early loading of *Mex67* to immature transcripts. Nuclear RNA surveillance mechanisms during the processing events and at the NPC ensure that only completely processed and correctly assembled mRNPs leave the nucleus while unprocessed or faulty mRNAs are retained and degraded.

The TREX Complex

The co-transcriptionally loaded TREX complex is essential for formation and export of the mRNP. TREX is conserved among eukaryotes and connects transcription elongation with mRNA maturation and export. TREX consists of the heterotetrameric transcription elongation complex THO and the export factors *Sub2* and *Yra1* (Strasser et al. 2002; Rougemaille et al. 2008). The recruitment of the TREX complex involves *Syf1*, which is a component of the *Prp19* splicing complex (Chanarat et al. 2011). THO is composed of *Hpr1*, *Mft1*, *Thp2*, and the eponymous *Tho2*. RNA/DNA-protein crosslinking experiments revealed that THO directly interacts with chromatin and RNA (Jimeno et al. 2002; Pena et al. 2012). The THO complex contributes to transcription elongation and genetic stability by preventing the formation of DNA:RNA hybrids known as R-loops and transcription-associated recombination (Huertas and Aguilera 2003; Jimeno and Aguilera 2010).

Recently, another THO complex component was identified as *Tex1*. It stably associates with the mRNA as an integral part of the THO complex (Jimeno et al. 2002; Hurt et al. 2004; Gewartowski et al. 2012; Pena et al. 2012). However, depletion of *Tex1* has no effect on THO complex assembly and binding to nucleic acids (Pena et al. 2012). Moreover, while the other THO complex components

show hyper-recombination and mRNA export defects when mutated, only mild effects are observed in a *TEX1* deletion strain (Luna et al. 2005).

THO binding to nascent mRNAs leads to the recruitment of *Yra1* and the DEAD-box RNA helicase *Sub2*, which is involved in early and late steps of spliceosome assembly (Jensen et al. 2001b; Strasser and Hurt 2001; Strasser et al. 2002; Hurt et al. 2004). Like observed for components of the THO complex, *Sub2* mutants show an elongation-dependent hyper-recombination phenotype and mRNA export defects (Chavez et al. 2000; Strasser et al. 2002; Garcia-Rubio et al. 2008). In higher eukaryotes, *Sub2* is part of the exon junction complex (EJC) that marks the exon-exon boundaries, however, this complex has not been identified in fungi yet (Abruzzi et al. 2004).

Interestingly, like *Sub2* and *Yra1*, the *Prp19* complex also binds to intronless pre-mRNAs suggesting a function that is not limited to spliceosome assembly and splicing (Lei and Silver 2002; Abruzzi et al. 2004; Chanarat et al. 2011). This is very similar to *Npl3*. This SR protein helps to recruit the splicing machinery onto intron containing transcripts (Kress et al. 2008), but is also present on intron free mRNAs (Lee et al. 1996; Krebber et al. 1999; Kim Guisbert et al. 2005). For *Sub2* an additional function was reported in inhibiting the transcription activity of RNA pol II toward the 3′ end, to prevent premature polyadenylation and mRNP release (Saguez et al. 2008).

A THO complex-dependent recruitment of the shuttling serine-arginine (SR) rich proteins *Gbp2* and *Hrb1* to the pre-mRNA has been shown earlier (Hacker and Krebber 2004; Hurt et al. 2004). Interestingly, their recruitment is further dependent on the splicing machinery, where both proteins play a crucial role in the quality control of splicing. Upon correct completion of splicing, both SR proteins bind to *Mex67* (Hackmann et al. 2014).

Earlier studies suggested that upon loading of *Sub2* onto the pre-mRNP *Yra1* binds and subsequently recruits *Mex67-Mtr2*, which is necessary for the export of the mature mRNP (Zenklusen et al. 2001; Stewart 2010). However, recent studies provide evidence for an alternative model in which the *Yra1* association occurs independently of *Sub2*, but via interaction with the 3′-end processing factor *Pcf11*. The protein binds to the phosphorylated CTD, recruits *Yra1* to the transcription elongation complex and the emerging mRNA (Johnson et al. 2009, 2011). *Yra1* itself contains a phospho-CTD-interacting domain, which also comprises an RNA recognition motif (RRM). Deletion of this domain leads to strong mRNA export defects (Stewart 2010; MacKellar and Greenleaf 2011). Strikingly, *Yra1* unlike the other *Mex67* interacting factors, does not shuttle with the mature mRNP to the cytoplasm, which leaves the question unanswered if *Mex67* upon *Yra1* loading contacts the mRNA directly or if it is transferred to another adapter protein after release of *Yra1*. Moreover, *Sub2* and *Mex67* share the binding site for *Yra1* so that their binding is mutually exclusive. Only upon *Sub2* release, *Mex67* can bind to *Yra1* (Strasser and Hurt 2001; Johnson et al. 2011).

Recently, the ATP-dependent RNA helicase *Dbp2* was introduced as a new mRNP remodeling factor that permits the recruitment of *Yra1* and *Nab2* to prevent premature 3′ end processing (Ma et al. 2013). In vivo crosslinking experiments

have shown that a decreased association of *Yra1*, *Nab2*, and *Mex67* to mRNA is detectable in *dbp2Δ* cells. Moreover, upon *Yra1* binding to the mRNA and physical interaction with *Dbp2*, *Yra1* finally inhibits the *Dbp2* unwinding activity, which might result in the *Dbp2* release from the mRNP. The *Yra1* induced release of *Dbp2* was suggested to represent a quality control step during transcription termination and 3′-end processing (Cloutier et al. 2012; Ma et al. 2013).

The TREX-2 and the SAGA Complexes

The *TREX-2* and the SAGA complexes cooperate to localize the expression of certain genes to the NPC for an efficient coupling of transcription, mRNP assembly and transport through the NPC (Stewart 2010). The *TREX-2* complex is localized in close proximity to the nuclear pore complex and assists the *TREX* complex in mRNA transcription and export. *TREX-2* consists of *Thp1*, *Cdc31*, *Sac3* and *Sus1*. Mutations of *TREX-2* complex factors show similar defects in transcription and mRNA export as observed for mutants of the *TREX* complex (Rodriguez-Navarro et al. 2004; Dieppois et al. 2006; Luthra et al. 2007; Jani et al. 2009). While *Sac3*, *Sus1* and *Cdc31* interact with the nucleoporin *Nup1* to connect the complex with the NPC, a linkage to the transcription process is provided by a physical interaction of *Thp1* and *Sus1* with the *S*pt-*A*da-*G*cn5 *a*cetyltransferase (SAGA) complex. Additionally, the nuclear basket associated protein Mlp1 binds to the promoter of active genes through interactions with SAGA subunits (Dieppois et al. 2006; Luthra et al. 2007).

The SAGA complex functions as a co-activator for transcription initiation by acetylating histones during transcription and thereby supporting the accessibility of transcription complexes to genomic DNA (Rodriguez-Navarro et al. 2004). Sus1 is part of both complexes, the *TREX-2* and the SAGA complex, and co-purifies with RNA pol II and the mRNA export factors *Yra1* and *Mex67* (Rodriguez-Navarro et al. 2004; Pascual-Garcia et al. 2008). *Sus1* might therefore couple transcription of SAGA-bound chromatin to the subsequent export of the transcripts by interaction with the NPC-attached *TREX-2* complex and *Mex67*. This tethering of certain actively transcribed genes (e.g., cell cycle regulated or less abundant transcripts) to the vicinity of the NPC allows an immediate export of such transcripts, which may compete with the export of highly and/or constitutively expressed mRNAs (Rodriguez-Navarro et al. 2004).

mRNA Export Receptors

The only known mRNA export receptors to date are *Mex67-Mtr2* and *Xpo1/Crm1*, although more might exist (Hieronymus and Silver 2003; Rodriguez-Navarro et al. 2004; Carmody and Wente 2009). The key factor for mRNA export is the

heterodimer *Mex67-Mtr2*, which works independently of the Ran GTPase system, required for general protein transport. As *Mex67* has a low affinity for direct binding to mRNA, the interaction seems to be mediated by several mRNA covering adaptor proteins (Strasser et al. 2002; Kohler and Hurt 2007). As already mentioned and to summarize at this point, so far the mRNA binding proteins *Hpr1* (of the THO-complex), *Yra1* (of the TREX-complex), *Npl3*, *Gbp2* and *Hrb1* (of the SR-protein family) and *Nab2* (involved in 3′-end processing) are known to interact with *Mex67* (Zenklusen et al. 2001; Gilbert and Guthrie 2004; Hobeika et al. 2007; Batisse et al. 2009; Iglesias et al. 2010; Hackmann et al. 2014). Interestingly, for the export of the large ribosomal subunit *Mex67* was suggested to directly contact the 5S rRNA as was shown in vitro experiments (Yao et al. 2007). However, adapter proteins might contribute to this binding in vivo.

The interaction domain necessary for the NPC binding of *Mex67* has been mapped to its C-terminus, which also represents the *Mtr2* binding domain. *Mtr2* in turn interacts with the phenylalanine/glycine (FG)-rich repeats of the nucleoporins (Nups). During translocation *Mtr2* exposes its Nup-binding sites, shielding the transport cargo from the hydrophobic meshwork of the inner channel and thereby allowing translocation (Kohler and Hurt 2007).

A second mRNA export receptor is *Crm1*/*Xpo1*, which is involved in the transport of very few mRNAs (Ohno et al. 2000; Kohler and Hurt 2007). *Crm1*/*Xpo1* is a karyopherin and uses the Ran GTPase system for export. It functions in the transport of nuclear export signal (NES) containing cargoes, including NES containing proteins associated with UsnRNAs, pre-ribosomal subunits and mRNAs (Fornerod et al. 1997). In yeast only slight mRNA export defects can be detected when *XPO1* is mutated (Hodge et al. 1999; Neville and Rosbash 1999). One interesting example of an *Xpo1*-transported mRNA is the *YRA1* transcript, which controls its own expression (Dong et al. 2007).

mRNA Export Adapters

mRNA export adapter proteins connect the receptors to the mRNA. While for the *Xpo1*-mediated export these are currently unknown, several factors were identified for *Mex67*. Of those, some adapter interactions are transient while other adapter proteins stay bound for an extended time and accompany the mRNA through the NPC into the cytoplasm where some of these proteins can persist until translation (Windgassen et al. 2004). Those that leave the mRNA already in the nucleus are *Yra1* and *Hpr1*, which interact with *Mex67* (Zenklusen et al. 2001; Hobeika et al. 2007). *Yra1* contacts *Mex67* upon release of *Sub2*, as they share the binding sites (Johnson et al. 2009; Iglesias et al. 2010). It is currently unclear if upon their dissociation *Mex67* binds to the mRNA or is transferred to other proteins.

Those adapter proteins that escort the mature transcript to the cytoplasm are *Nab2* and the three SR-proteins, *Npl3*, *Gbp2* and *Hrb1*. As already pointed out, all three yeast SR-proteins are recruited to the pre-mRNA co-transcriptionally,

however, *Npl3* is recruited early by RNA pol II prior to splicing and *Gbp2* and *Hrb1* to a later time point via the THO complex and in dependence of splicing (Lei et al. 2001; Hacker and Krebber 2004; Hurt et al. 2004; Hackmann et al. 2014). *Npl3* contains two typical RNA recognition motifs (RRM) for RNA binding, an N-terminal domain with several APQE (alanine, proline, glutamine, and glutamate) repeats of unknown function and a C-terminal domain with several SR (serine, arginine) and RGG (arginine, glycine, glycine) repeats, termed the SR-domain. Interestingly, the deletion of *NPL3* in certain backgrounds of *Saccharomyces cerevisiae*, such as *BY4743*, does not cause visible mRNA export defects, revealing that other factors can compensate for the function of *Npl3* in mRNA transport (Hackmann et al. 2011). The growth defects visible in this background rather result from defects in ribosomal subunit joining important for translation initiation (Baierlein et al. 2013). In yeast strains in which *NPL3* is essential such as *S288C* or *W303*, mutations in its gene lead to mRNA export defects (Lee et al. 1996; Krebber et al. 1999). Moreover, mutations in both RRMs of *NPL3* do not only lead to mRNA but also to pre-60S export defects (Stage-Zimmermann et al. 2000; Hackmann et al. 2011). *Npl3* physically interacts with *Mex67* and this interaction is not dependent on RNA (Gilbert and Guthrie 2004; Hackmann et al. 2014). Strikingly, *Npl3* can also interact directly with FG-rich elements of the nucleoporins, which might be important for the nuclear export of large ribosomal subunits, because the association of *Mex67* to pre-60S ribosomal subunits is independent of *Npl3* (Hackmann et al. 2011). However, it might also suggest that the protein is more than just an adapter for *Mex67*, but rather actively contributes to the shielding of the transport particle from the hydrophobic interior of the NPC, which might also be conceivable for the transport of mRNAs.

Two other nuclear functions have been reported for *Npl3*: (a) *Npl3* is involved in the co-transcriptional assembly of the early spliceosome on the nascent pre-mRNAs and it interacts genetically and physically in the presence of RNA with U1 and U2 snRNP splicing factors (Kress et al. 2008). Moreover, deletion of *NPL3* leads to an accumulation of intron containing pre-mRNAs (Kress et al. 2008). (b) Furthermore, *Npl3* acts as an anti-terminator by competing for RNA binding with the 3′-end processing machinery and in this way prevents polyadenylation at weak polyadenylation signal sequences (Bucheli and Buratowski 2005; Bucheli et al. 2007; Dermody et al. 2008). Although *Npl3* has several nuclear functions and most of the protein is localized to the nucleus, an essential cytoplasmic function has been discovered recently in which *Npl3* mediates the joining of the ribosomal subunits during translation initiation for which it needs to form dimers (Baierlein et al. 2013). Furthermore, *Npl3* is the target of several different posttranslational modifications, which have an impact on several functions and protein-protein interactions (Inoue et al. 2000; Gilbert et al. 2001; McBride et al. 2005).

Gbp2 and *Hrb1* are recruited during transcription elongation to pre-mRNAs by the THO complex (Hacker and Krebber 2004; Hurt et al. 2004). Their recruitment further depends on splicing and their initial binding to bulk mRNA is tightened on intron containing transcripts by the splicing machinery (Hackmann et al. 2014). Recently, the Tollervey lab analyzed the composition of different RNP complexes

and found increased interactions of *Gbp2* with spliced transcripts at their 5′-ends (Tuck and Tollervey 2013). In contrast to that, *Mex67* was detected all over the mRNA sequences, revealing binding to the nascent transcript at various points (Tuck and Tollervey 2013). *Gbp2* and *Hrb1* share a significant homology in their amino acid sequence (47 % identity) (Hacker and Krebber 2004). Both contain three RRM motifs and an N-terminal SR/RGG domain. Strikingly, deletions of *Gbp2* and *Hrb1*, show no mRNA export defects, but rather the opposite, an increased slippage of unspliced transcripts into the cytoplasm, which suggests a function in retaining unspliced pre-mRNAs in the nucleus, important for mRNA surveillance (Hackmann et al. 2014). In fact, *Gbp2* and *Hrb1* stabilize the TRAMP complex association with the transcript and channel false transcripts into degradation. Upon TRAMP complex release from correct RNAs, *Gbp2* and *Hrb1* recruit *Mex67* for export (Hackmann et al. 2014).

After 3′-end processing and polyadenylation *Nab2* associates with the nascent poly(A) tail. *Nab2* is required for the trimming of the poly(A) tail and for nuclear mRNA export via association of *Mex67* (Hector et al. 2002). *Yra1* participates in 3′-processing as mutations in *Yra1* that are otherwise lethal, can be suppressed by overexpression of *Mex67* or *Nab2* (Anderson et al. 1993; Hector et al. 2002; Iglesias et al. 2010). The N-terminal domain of Nab2 physically interacts with the nucleoporin Mlp1 (Fasken et al. 2008). Therefore, it was suggested that Nab2 may help to concentrate mature mRNAs at the nuclear face of the NPC for nuclear export (Soucek et al. 2012).

In summary, different mRNA adapter proteins are recruited over the course of mRNA maturation, some of which dissociate prior to export, some of which remain bound to the mRNA during translocation. However, all have in common that they recruit *Mex67-Mtr2*, which might function as flags for the properly matured mRNA that are recognized at the NPC by Mlp1 (Hackmann et al. 2014). During transit through the NPC it covers, likely together with other molecules, the charged backbone of the mRNA from the hydrophobic interior of the NPC.

The Late Phase: The Nuclear Pore Complex and mRNA Translocation

Once the mRNAs are matured and covered with *Mex67-Mtr2* molecules they are ready for their passage through the NPC. These eight-fold symmetrical complexes with a molecular mass of ~50 MDa are highly conserved and comprised of ~30 different proteins, termed nucleoporins (Nups) (Strambio-De-Castillia et al. 2010a, b). NPCs are embedded into the nuclear envelope with two coaxial rings positioned coplanar with the inner and outer membranes (Fig. 4.2). This central structure extends into the cytoplasm with eight cytoplasmic filaments that are connected with the cytoskeleton to alleviate the way toward protein synthesis. The nuclear basket on the other side channels incoming cargo toward the nuclear

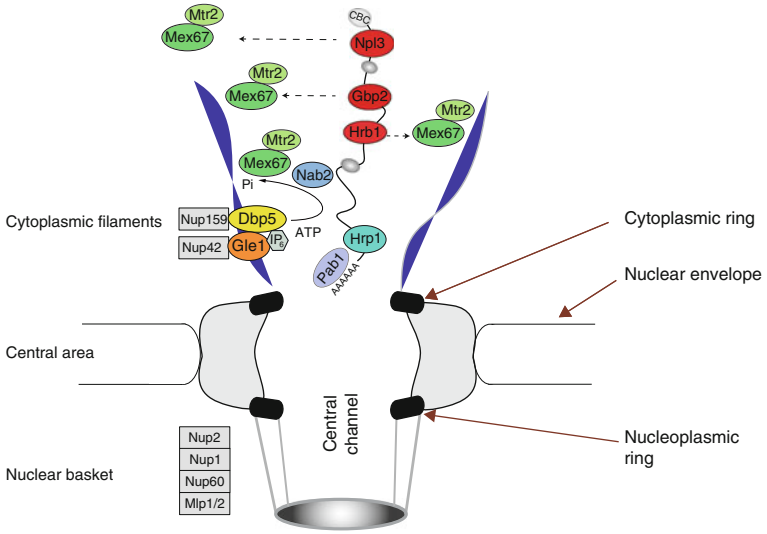


Fig. 4.2 Release of mRNA transport factors from the mRNP after export through the Nuclear Pore Complex (NPC). Upon arrival of the mRNP at the cytoplasmic side of the NPC *Dbp5* displaces *Mex67-Mtr2* and *Nab2* from the mRNP in an ATP-dependent step, leading to the directionality of the transport event

interior (Pante and Aebi 1995; Fahrenkrog and Aebi 2003; Strambio-De-Castillia et al. 2010a, b). Current views reflect the NPCs not as isolated complexes through which transport occurs, but rather as areas for gated, selectively promoted gene expression, platforms for macromolecular assembly and funnels for efficient translation (Strambio-De-Castillia et al. 2010a, b). The central channel has an approximate diameter of 35 nm and is filled with flexible filamentous FG (phenylalanine-glycine) Nups that prevent macromolecules from entering. Cargoes bound to transport receptors can overcome this hydrophobic barrier, which due to its flexible nature, accommodates the transport of differently sized cargoes. FG Nups are characterized by regions of multiple FG-repeats separated by hydrophilic spacer sequences of 5–30 amino acid residues. In *S. cerevisiae* around 160 individual FG Nups edge the transport channel in each NPC (Rout et al. 2000; Alber et al. 2007). The NPC contains a central area, a nuclear basket and cytoplasmic filaments (Fig. 4.2). The central area is composed of central FG Nups, inner ring- and outer ring Nups, linker Nups and transmembrane ring Nups (Strambio-De-Castillia et al. 2010a, b). One of the central FG Nups in *S. cerevisiae* is *Nup116* (a Nup only found in fungi; Sampathkumar et al. 2012), *Nup100*, *Nsp1*, *Nup57* and *Nup49* (Aitchison and Rout 2012). Interestingly, *Nup116* interacts with the NPC-associated protein *Gle2* (Rae1 in *S. pombe)*, which was suggested to play a role in mRNA export as mutants show mRNA export defects and *Gle2* was found in a complex with *Mex67*. Moreover, *Gle2* mutants show defects in the NPC structure (Wente and Blobel 1993; Murphy et al. 1996; Bailer et al. 1998; Pritchard et al. 1999).

The nuclear basket is made of *Nup1*, *Nup2*, and *Nup60* and two myosin-like proteins *Mlp1* and *Mlp2*, which form a filamentous structure that reaches into the nuclear interior (Galy et al. 2004, Strambio-De-Castillia et al. 2010a, b). *Mlp1* and *Mlp2* play a key role in the entry of mRNPs into the NPC (Green et al. 2003; Galy et al. 2004). They interact with the C-terminal part of *Mex67* and *Mex67*-interacting shuttling RNA binding proteins such as *Nab2*, *Npl3*, *Gbp2*, and *Hrb1* (Green et al. 2003; Fasken et al. 2008; Hackmann et al. 2014). As their deletion leads to the slippage of intron-containing transcripts into the cytoplasm, a key role in nuclear quality control had been proposed for *Mlp1* and *Mlp2* (Galy et al. 2004; Hackmann et al. 2014). As nuclear gate keepers, the Mlp proteins might monitor the completed mRNA maturation by checking for proper *Mex67* association (Hackmann et al. 2014). *Sac3* is an NPC associated factor, which is part of the TREX-2 multiprotein complex that interacts with *Nup1*. Its association is stabilized via *Sus1* and *Cdc31* (Garcia-Olivier et al. 2012). *Sac3* supports anchoring of mature *Mex67*-bound mRNPs to the nuclear entry side of the NPC (Lei et al. 2003).

The cytoplasmic filaments of the NPC are composed of *Nup159* and *Nup42* and bind the DEAD-box RNA helicase *Dbp5*, which by its remodeling activity is the driving force of mRNA export and leads to directionality (Snay-Hodge et al. 1998; Tseng et al. 1998; Strambio-De-Castillia et al. 2010a, b; Aitchison and Rout 2012; Tieg and Krebber 2013) (Fig. 4.2). When the transported mRNP reaches the cytoplasmic side of the NPC, it contacts *Dbp5* and its co-factors *Gle1* and inositol hexakisphosphate (IP_6) that stimulate the ATPase activity of *Dbp5* (Alcazar-Roman et al. 2006; Weirich et al. 2006). The consequent *Dbp5* dependent mRNP remodeling was suggested to first link the RNA leading to the displacement of *Mex67* and *Nab2* and to subsequently release of the remaining mRNP into the cytoplasm, preventing a back-slippage of the mRNP into the nucleus (Tseng et al. 1998; Tran et al. 2007) (Fig. 4.2). This whole process occurs in several steps: *Dbp5* in its open conformation binds ATP, leading to a cooperative binding of Gle1-IP_6 and the exported mRNP. Binding causes a conformational change, leading to ATP hydrolysis. The transition to the ADP bound form leads to the specific displacement of mRNP bound proteins and to the dissociation of mRNA, allowing an interaction of *Dbp5* with *Nup159*. This interaction induces a conformational change resulting in the release of the ADP and the rebinding of ATP (Hodge et al. 2011; Noble et al. 2011; Tieg and Krebber 2013).

Interestingly, *Nup159* is specifically required for the export of mRNAs, as temperature sensitive mutants show strong mRNA export defects while the import of NLS containing proteins is unaffected (Gorsch et al. 1995; Del Priore et al. 1997). Another *Dbp5* associated protein of unknown function is *Gfd1* (*g*ood *f*or *Dbp5*), which in high copy suppresses temperature sensitive mutants of *Dbp5* (Zheng et al. 2010). Interestingly, besides their function in mRNA export, *Dbp5* and *Gle1* have been shown to function in translation termination together with the eukaryotic release factor 1 (eRF1) (Bolger et al. 2008). This involvement of mRNA export factors in other essential subsequent processes, such as *Dbp5* in

translation termination, *Npl3* in translation initiation and *Hrp1* in NMD, again emphasizes the intimate coupling of basic functions in gene expression and reveals the high efficiency in nature of using one protein for several cellular functions.

mRNA Surveillance and Quality Control in the Nucleus

The synthesis of pre-mRNAs and their processing to mature mRNPs are not error free processes. Consequently, the cell has evolved different mechanisms to prevent the expression of false transcripts. Similar to the intimate linkage of transcription and export, the nuclear quality control is also tightly connected with pre-mRNA maturation and occurs at several steps. Co-transcriptional degradation processes are mainly executed by the nuclear exosome and co-factors like the Nrd1-Nab3-Sen1 complex for degradation of premature cryptic transcripts, the Ccr4-Not complex that might be a scaffold for the assembly of factors involved in ubiquitination and deadenylation, and the TRAMP complex, which marks RNAs with short oligo(A) tails to initiate their degradation (Fasken and Corbett 2009; Houseley and Tollervey 2009; Collart and Panasenko 2012; Porrua and Libri 2013).

The exosome is an evolutionary conserved nine-subunit complex including Dis3/Rrp44 that exhibits exo- and endoribonuclease activity (see Chap. 7 for more details). The exclusively nuclear exosome component *Rrp6* contains additional exonuclease activity. The main co-factor of the exosome is the *Tr*f-*A*ir-*M*tr4 *p*olyadenylation complex (TRAMP), which exists in two forms that recognizes different RNA substrates (Houseley and Tollervey 2009; Kong et al. 2013; Tuck and Tollervey 2013). TRAMP5 consists of a noncanonical poly(A) polymerase (*Trf5*), a zinc-knuckle RNA binding protein (*Air1*) and the DExH-box RNA helicase *Mtr4* and seems to be mostly restricted to nucleolar rRNA processing. The TRAMP4-complex (*Trf4*, *Air2*, and *Mtr4*) is the major co-factor that mediates the nuclear quality control by recognizing aberrant RNAs and directing them for degradation to the nuclear exosome (San Paolo et al. 2009; Callahan and Butler 2010). The mechanism by which RNAs are recognized as aberrant is not fully understood. However, it has been shown that the zinc-knuckle Air proteins provide the initial contact of the TRAMP complex to the RNA and modulate the Trf4/5 polymerase activity (San Paolo et al. 2009; Hamill et al. 2010). In contrast to the general poly(A) polymerase *Pap1*, which adds 70–90 nucleotides long poly(A) tails in *S. cerevisiae*, the polymerases *Trf4*/*Pap2* and *Trf5* add only short poly(A) tails of 10 or less nucleotides to the 3′-end of an RNA, which could act as a label for degradation. The RNA helicase Mtr4 controls the length of the oligo(A) tail and unwinds secondary structures, leading to an unstructured 3′-end that can enter the exosome (Jia et al. 2011).

As far as it is known, defects in early transcription are monitored and the defective RNAs are eliminated by the immediate recruitment of the exosome (Hilleren et al. 2001; Jensen et al. 2003). How abnormal pre-mRNAs are released from the transcription apparatus and whether co-factors are involved in this

process is currently unclear. However, the SR-protein *Npl3*, which is recruited early during transcription, might be the most 5′ located protein that receives a *Mex67* molecule, which might indicate proper 5′ maturation (Lei et al. 2001; Hackmann et al. 2014). A recent publication in which the *cr*osslinking and *a*nalysis of *c*DNA (CRAC) technique was applied, revealed that the TRAMP4 complex factors *Trf4* and *Mtr4* bind the transcripts in their 5′ regions, close to the transcription start site, which is crucial for an early RNA surveillance (Granneman et al. 2009; Tuck and Tollervey 2013). The experiments further showed a high binding density of these proteins on intron sequences, which are in yeast usually very close to the 5′ end of the transcripts. Moreover, an interaction with several splicing factors was shown for *Trf4* and *Mtr4*, which support the idea of a TRAMP and exosome-dependent surveillance mechanism of pre-mRNA splicing (Kong et al. 2013; Tuck and Tollervey 2013). In fact, key factors in nuclear surveillance of spliced transcripts are the SR proteins *Gbp2* and *Hrb1* because they connect the TRAMP complex to spliced transcripts and if splicing is not properly executed, initiate their degradation. On correct RNAs they instead recruit *Mex67* to signal export competence (Hackmann et al. 2014). Another nuclear quality control protein is the pre-mRNA *re*tention and *s*plicing (RES)-complex protein *Pml1* that was suggested to contact intron containing mRNAs prior to *Mlp1* (Dziembowski et al. 2004; Palancade et al. 2005). However, the function of *Pml1* in this process is rather nebulous.

Along the road, *Yra1* and *Nab2* associate with the transcripts and might control 3′-end processing events. On correctly processed mRNAs the association of *Mex67* molecules close to their 3′ ends mark them for export. Interestingly, the CRAC assay revealed a second place of *Nab2* association close to the 5′ end of transcripts, suggesting that the protein not only acts in 3′-end processing events (Tuck and Tollervey 2013). Additionally, all defects in transcription, splicing and 3′-end processing might lead to a delay or failure of *Pap1* mediated polyadenylation that favours oligo-adenylation by TRAMP and subsequent degradation by the exosome (Tutucci and Stutz 2011).

Prior to export, mRNPs are surveyed by a quality control checkpoint directly situated at the nuclear basket of the NPC. *Mlp1* and *Mlp2*, both anchored by *Nup60* block the export of intron-containing mRNAs and aberrantly assembled mRNPs at the NPC. Upon deletion of their genes all three of them show leakage of unspliced transcripts into the cytoplasm (Galy et al. 2004; Palancade et al. 2005). *Pml39* (*Rsm1* in *S. pombe*) is an NPC-associated factor that contacts *Nup84* and the Mlp proteins that show leakage of unspliced pre-mRNA into the cytoplasm when deleted (Palancade et al. 2005). The integral inner nuclear membrane protein *Esc1* interacts with *Mlp1* and *Mlp2*. *Esc1* functions in maintaining the correct composition of the nuclear basket and therefore might be responsible for proper positioning of the NPC-associated surveillance factors (Lewis et al. 2007; Niepel et al. 2013). The cytoplasmic appearance of unspliced transcripts can also be observed in mutants of *Ulp1*. The SUMO protease *Ulp1* was suggested to contribute to intron-containing mRNA retention at the NPC by de-sumoylation of pre-mRNA associated proteins that were marked for degradation. *Ulp1* is localized to the NPC

by association with *Nup2*, *Esc1*, *Nup60* and *Mlp1* (Lewis et al. 2007). Finally, *Swt1* is an RNA endoribonuclease, which is transiently recruited to NPCs. The inactivation of its endonuclease activity leads to leakage of intron mRNAs to the cytoplasm. Possibly, the endonucleolytic cleavage of false transcripts by *Swt1* is necessary for the degradation of such messages (Skruzny et al. 2009).

An early model of how defective transcripts are detected at the NPC suggests that *Mlp1* might directly contact an intron-associated factor to recognize unspliced mRNAs, as the branch point binding protein *Msl5* was shown to interact with *Mlp1* in an RNA-dependent manner (Galy et al. 2004). On the other side, *Mlp1* also reduces chromatin crowding and might therefore contribute to gene gating (Niepel et al. 2013). However, *Mlp1* might alternatively rather be a detector of proper *Mex67* association. Like a ticket controller, it might survey the transcript for proper *Mex67* coverage (Hackmann et al. 2014). Indeed, the CRAC technique revealed an even distribution of *Mex67* over the entire transcript length (Tuck and Tollervey 2013), supporting the stepwise recruitment of *Mex67* upon completion of individual processing steps.

mRNA Export and its Regulation Via Post-Translational Protein Modifications

Gene expression is regulated by a complex network, which coordinates mRNA synthesis, processing, export, translation, and several layers of quality control. These events are affected by phosphorylation, methylation, and ubiquitination of mRNA-associated factors. Foremost these post-transcriptional modifications influence protein–protein or protein-RNA interactions. Differential phosphorylation of the CTD of the RNA pol II leads to the recruitment of certain proteins at specific time points to guarantee a smooth maturation of the transcripts, but also the loaded proteins are themselves modified. The following examples will describe a few cases in which the impact of the modifications is quite well understood.

The SR protein *Npl3* is phosphorylated and methylated. One phosphorylation site in the SR motif closest to the C-terminus is targeted by the cytoplasmic kinase *Sky1* (Gilbert et al. 2001). Phosphorylated *Npl3* leads to a reduced RNA binding affinity and an increased reimport into the nucleus mediated by the SR protein specific karyopherin *Mtr10* (Gilbert et al. 2001). Interestingly, the reduced RNA binding affinity seems to be important for its nuclear association with the mRNA, rather than for its RNA dissociation in the cytoplasm, and although the deletion strain of the cytoplasmic kinase Sky1 shows an increased poly(A)$^+$RNA binding phenotype of *Npl3*, its dissociation from mRNAs engaged in translation is not affected (Windgassen et al. 2004). It was rather shown that the dissociation of *Npl3* from polysomal mRNAs requires *Mtr10* (Windgassen et al. 2004). Upon loading of *Npl3* onto the emerging pre-mRNA in the nucleus, *Npl3* is phosphorylated during later steps of transcription, which supports the binding of Rna15 to the polyadenylation signal to initiate 3′-end cleavage and polyadenylation.

Finally, *Npl3* is dephosphorylated by the exclusively nuclear phosphatase *Glc7*, which leads to the dissociation of the 3′-end processing factors and importantly promotes loading of the mRNA export receptor *Mex67* (Gilbert and Guthrie 2004).

Additionally, it was shown that several RGG motifs of *Npl3* are methylated by the methyltransferase Hmt1, which decreases the interactions of *Npl3* with the CBC, *Tho2*, and with itself (McBride et al. 2005; Erce et al. 2013). These results provide evidence that upon *Npl3* methylation, interactions with processing and transcription elongation factors are loosened to promote the release of the export competent mRNP from the transcription apparatus, which might support an interaction of *Npl3* with the export receptor *Mex67* (McBride et al. 2005). Interestingly, unmethylated *Npl3* does not seem to shuttle to the cytoplasm anymore, however, a knock-out of *HMT1* is viable and shows no bulk mRNA export defects (Shen et al. 1998). Less is known about the two other SR proteins *Gbp2* and *Hrb1*, but first evidence exists that they might like *Npl3* undergo methylation and phosphorylation (Windgassen and Krebber 2003; Erce et al. 2013).

The THO complex subunit *Hpr1* has been shown to be ubiquitinated in its C-terminal region during transcription elongation (Hobeika et al. 2009; Gewartowski et al. 2012). The ubiquitinated protein is recognized and bound by the C-terminal domain of the export receptor *Mex67*. A block in *Hpr1* ubiquitination results in a decrease of co-transcriptional recruitment of *Mex67* to the mRNA. Structural studies revealed that ubiquitinated *Hpr1* and the NPC subunits may bind to the *Mex67* ubiquitin associated domain in a mutually exclusive manner (Hobeika et al. 2007; Iglesias et al. 2010).

Moreover, *Yra1* is ubiquitinated by the E3 ubiquitin ligase *Tom1*. This ubiquitination leads to the dissociation of *Yra1* from the mRNP and coincides with the delivery of *Mex67* to *Nab2* (Iglesias et al. 2010). Interestingly, *Tom1* interacts genetically with *Rrp6* and *Mlp2* and loss of perinuclear Mlp proteins suppress the growth defects of *Tom1* and *Yra1* ubiquitination mutants, suggesting that *Tom1*-mediated dissociation of *Yra1* from *Nab2*-bound mRNAs is part of a surveillance mechanism at the pore, ensuring the export of matured mRNPs (Iglesias et al. 2010).

mRNA Export During Cellular Stress

Environmental changes like heat shock, osmotic and oxidative stress, or nutrient starvation needs a cellular adaption by a rapid alteration in global gene expression patterns. Under stress conditions, the nuclear export of regular mRNAs is blocked while stress-specific mRNAs are rapidly exported and translated in the cytoplasm (Saavedra et al. 1996, 1997). This nuclear mRNA retention might be caused by the dissociation of *Mex67* adapter proteins such as *Npl3* and *Nab2* from bulk mRNAs (Krebber et al. 1999; Carmody et al. 2010). For *Nab2* it was shown that the MAP kinase *Slt2* phosphorylates *Nab2*, leading to a decreased binding of *Mex67* and an accumulation of *Nab2*, *Yra1* and *Mlp1* in nuclear foci (Carmody et al. 2010). The mRNA export block in the nucleus also coincides with the release of the

NPC-associated *Gle2* into the cytoplasm upon heat and ethanol stress. Furthermore, the 3′-end processing factor *Hrp1* shifts from the nuclear localization at steady state to a cytoplasmic enrichment during osmotic stress, and shifts back to the nucleus upon removal of the stress condition (Henry et al. 2003; Tutucci and Stutz 2011). Interestingly, *Npl3*, *Yra1* and THO complex factor mutants which regularly lead to poly(A)$^+$ RNA export defects under normal conditions do not inhibit the export of heat shock mRNAs, while *Mex67-5* does (Rollenhagen et al. 2007; Gewartowski et al. 2012). It is currently unclear how *Mex67* mediates the transport of stress specific mRNAs and if other proteins are involved.

References

Abruzzi KC, Lacadie S, Rosbash M (2004) Biochemical analysis of TREX complex recruitment to intronless and intron-containing yeast genes. EMBO J 23:2620–2631

Ahn SH, Kim M, Buratowski S (2004) Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3′ end processing. Mol Cell 13:67–76

Aitchison JD, Rout MP (2012) The yeast nuclear pore complex and transport through it. Genetics 190:855–883

Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, Devos D et al (2007) The molecular architecture of the nuclear pore complex. Nature 450:695–701

Alcazar-Roman AR, Tran EJ, Guo S, Wente SR (2006) Inositol hexakisphosphate and Gle1 activate the DEAD-box protein Dbp5 for nuclear mRNA export. Nat Cell Biol 8:711–716

Anderson JT, Wilson SM, Datar KV, Swanson MS (1993) NAB2: a yeast nuclear polyadenylated RNA-binding protein essential for cell viability. Mol Cell Biol 13:2730–2741

Babour A, Dargemont C, Stutz F (2012) Ubiquitin and assembly of export competent mRNP. Biochim Biophys Acta 1819:521–530

Baierlein C, Hackmann A, Gross T, Henker L, Hinz F, and Krebber H (2013) Monosome formation during translation initiation requires the serine/arginine-rich protein Npl3. Mol Cell Biol 33(24):4811–4823

Bailer SM, Siniossoglou S, Podtelejnikov A, Hellwig A, Mann M, Hurt E (1998) Nup116p and nup100p are interchangeable through a conserved motif which constitutes a docking site for the mRNA transport factor *gle2p*. EMBO J 17:1107–1119

Barilla D, Lee BA, Proudfoot NJ (2001) Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA 98:445–450

Barnwal RP, Lee SD, Moore C, Varani G (2012) Structural and biochemical analysis of the assembly and function of the yeast pre-mRNA 3′ end processing complex CF I. Proc Natl Acad Sci U S A 109:21342–21347

Bataille AR, Jeronimo C, Jacques PE, Laramee L, Fortin ME, Forest A et al (2012) A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. Mol Cell 45:158–170

Batisse J, Batisse C, Budd A, Bottcher B, Hurt E (2009) Purification of nuclear poly(A)-binding protein Nab2 reveals association with the yeast transcriptome and a messenger ribonucleoprotein core structure. J Biol Chem 284:34911–34917

Bentley DL (2005) Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. Curr Opin Cell Biol 17:251–256

Bolger TA, Folkmann AW, Tran EJ, Wente SR (2008) The mRNA export factor Gle1 and inositol hexakisphosphate regulate distinct stages of translation. Cell 134:624–633

Bucheli ME, Buratowski S (2005) Npl3 is an antagonist of mRNA 3′ end formation by RNA polymerase II. EMBO J 24:2150–2160

Bucheli ME, He X, Kaplan CD, Moore CL, Buratowski S (2007) Polyadenylation site choice in yeast is affected by competition between Npl3 and polyadenylation factor CFI. RNA 13:1756–1764

Buratowski S (2009) Progression through the RNA polymerase II CTD cycle. Mol Cell 36:541–546

Callahan KP, Butler JS (2010) TRAMP complex enhances RNA degradation by the nuclear exosome component Rrp6. J Biol Chem 285:3540–3547

Carmody SR, Wente SR (2009) mRNA nuclear export at a glance. J Cell Sci 122:1933–1937

Carmody SR, Tran EJ, Apponi LH, Corbett AH, Wente SR (2010) The mitogen-activated protein kinase Slt2 regulates nuclear retention of non-heat shock mRNAs during heat shock-induced stress. Mol Cell Biol 30:5168–5179

Chanarat S, Seizl M, Strasser K (2011) The Prp19 complex is a novel transcription elongation factor required for TREX occupancy at transcribed genes. Genes Dev 25:1147–1158

Chanarat S, Burkert-Kautzsch C, Meinel DM, Strasser K (2012) Prp19C and TREX: interacting to promote transcription elongation and mRNA export. Transcription 3:8–12

Chavez S, Beilharz T, Rondon AG, Erdjument-Bromage H, Tempst P, Svejstrup JQ et al (2000) A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in *Saccharomyces cerevisiae*. EMBO J 19:5824–5834

Chen HC, Cheng SC (2012) Functional roles of protein splicing factors. Biosci Rep 32:345–359

Cloutier SC, Ma WK, Nguyen LT, Tran EJ (2012) The DEAD-box RNA helicase Dbp2 connects RNA quality control with repression of aberrant transcription. J Biol Chem 287:26155–26166

Collart MA, Panasenko OO (2012) The Ccr4–not complex. Gene 492:42–53

Colot HV, Stutz F, Rosbash M (1996) The yeast splicing factor Mud13p is a commitment complex component and corresponds to CBP20, the small subunit of the nuclear cap-binding complex. Genes Dev 10:1699–1708

Del Priore V, Heath CC, Snay CA, MacMillan AG, Dagher S, Cole CN (1997) A structure/function analysis of Rat7p/Nup159p, an essential nucleoporin of *Saccharomyces cerevisiae*. J Cell Sci 110:2897–2999

Dermody JL, Dreyfuss JM, Villen J, Ogundipe B, Gygi SP, Park PJ et al (2008) Unphosphorylated SR-like protein Npl3 stimulates RNA polymerase II elongation. PLoS ONE 3:e3273

Dieppois G, Iglesias N, Stutz F (2006) Cotranscriptional recruitment to the mRNA export receptor *Mex67p* contributes to nuclear pore anchoring of activated genes. Mol Cell Biol 26:7858–7870

Dong S, Li C, Zenklusen D, Singer RH, Jacobson A, He F (2007) YRA1 autoregulation requires nuclear export and cytoplasmic Edc3p-mediated degradation of its pre-mRNA. Mol Cell 25:559–573

Dunn EF, Hammell CM, Hodge CA, Cole CN (2005) Yeast poly(A)-binding protein, Pab1, and PAN, a poly(A) nuclease complex recruited by Pab1, connect mRNA biogenesis to export. Genes Dev 19:90–103

Dziembowski A, Ventura AP, Rutz B, Caspary F, Faux C, Halgand F et al (2004) Proteomic analysis identifies a new complex required for nuclear pre-mRNA retention and splicing. EMBO J 23:4847–4856

Erce MA, Abeygunawardena D, Low JK, Hart-Smith G, Wilkins MR (2013) Interactions affected by arginine methylation in the yeast protein-protein interaction network. Mol Cell Proteomics 12:3184–3198

Fahrenkrog B, Aebi U (2003) The nuclear pore complex: nucleocytoplasmic transport and beyond. Nat Rev Mol Cell Biol 4:757–766

Fasken MB, Corbett AH (2009) Mechanisms of nuclear mRNA quality control. RNA Biol 6:237–241

Fasken MB, Stewart M, Corbett AH (2008) Functional significance of the interaction between the mRNA-binding protein, Nab2, and the nuclear pore-associated protein, Mlp1, in mRNA export. J Biol Chem 283:27130–27143

Fornerod M, Ohno M, Yoshida M, Mattaj IW (1997) CRM1 is an export receptor for leucine-rich nuclear export signals [see comments]. Cell 90:1051–1060

Galy V, Gadal O, Fromont-Racine M, Romano A, Jacquier A, Nehrbass U (2004) Nuclear retention of un-spliced mRNAs in yeast is mediated by perinuclear Mlp1. Cell 116:63–73

Garcia-Olivier E, García-Molinero V, Rodríguez-Navarro S (2012) mRNA export and gene expression: the SAGA-TREX-2 connection. Biochim Biophys Acta 1819:555–565

Garcia-Rubio M, Chavez S, Huertas P, Tous C, Jimeno S, Luna R et al (2008) Different physiological relevance of yeast THO/TREX subunits in gene expression and genome integrity. Mol Genet Genomics 279:123–132

Gewartowski K, Cuéllar J, Dziembowski A, Valpuesta JM (2012) The yeast THO complex forms a 5-subunit assembly that directly interacts with active chromatin. Bioarchitecture 2(4):134–137

Gilbert W, Guthrie C (2004) The glc7p nuclear phosphatase promotes mRNA export by facilitating association of *Mex67p* with mRNA. Mol Cell 13:201–212

Gilbert W, Siebel CW, Guthrie C (2001) Phosphorylation by Sky1p promotes Npl3p shuttling and mRNA dissociation. RNA 7:302–313

Gonzáles CI, Ruiz-Echevarria MJ, Vasudevan S, Henry MF, Peltz SW (2000) The yeast hnRNP-like protein Hrp1/Nab4 marks a transcript for nonsense-mediated mRNA decay. Mol Cell 5:489–499

Gornemann J, Kotovic KM, Hujer K, Neugebauer KM (2005) Co-transcriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. Mol Cell 19:53–63

Gorsch LC, Dockendorff TC, Cole CN (1995) A conditional allele of the novel repeat-containing yeast nucleoporin RAT7/NUP159 causes both rapid cessation of mRNA export and reversible clustering of nuclear pore complexes. J Cell Biol 129:939–955

Granneman S, Kudla G, Petfalski E, Tollervey D (2009) Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. Proc Natl Acad Sci USA 106:9613–9618

Green DM, Marfatia KA, Crafton EB, Zhang X, Cheng X, Corbett AH (2002) Nab2p Is Required for Poly(A) RNA Export in *Saccharomyces cerevisiae* and Is Regulated by Arginine Methylation via Hmt1p. J Biol Chem 277:7752–7760

Green DM, Johnson CP, Hagan H, Corbett AH (2003) The C-terminal domain of myosin-like protein 1 (Mlp1p) is a docking site for heterogeneous nuclear ribonucleoproteins that are required for mRNA export. Proc Natl Acad Sci USA 100:1010–1015

Gwizdek C, Iglesias N, Rodriguez MS, Ossareh-Nazari B, Hobeika M, Divita G et al (2006) Ubiquitin-associated domain of *Mex67* synchronizes recruitment of the mRNA export machinery with transcription. Proc Natl Acad Sci USA 103:16376–16381

Hacker S, Krebber H (2004) Differential export requirements for shuttling serine/arginine-type mRNA-binding proteins. J Biol Chem 279:5049–5052

Hackmann A, Gross T, Baierlein C, Krebber H (2011) The mRNA export factor Npl3 mediates the nuclear export of large ribosomal subunits. EMBO Rep 12:1024–1031

Hackmann A, Wu A, Schneider UM, Meyer K, Jung K, Krebber H (2014) Quality control of spliced mRNAs requires the shuttling SR-proteins Gbp2 and Hrb1. Nat Commun 5(3123):1–14

Hamill S, Wolin SL, Reinisch KM (2010) Structure and function of the polymerase core of TRAMP, a RNA surveillance complex. Proc Natl Acad Sci USA 107:15045–15050

Hector RE, Nykamp KR, Dheur S, Anderson JT, Non PJ, Urbinati CR et al (2002) Dual requirement for yeast hnRNP Nab2p in mRNA poly(A) tail length control and nuclear export. EMBO J 21:1800–1810

Heidemann M, Hintermair C, Voss K, Eick D (2013) Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. Biochim Biophys Acta 1829:55–62

Henry MF, Mandel D, Routson V, Henry PA (2003) The yeast hnRNP-like protein Hrp1/Nab4 sccumulates in the cytoplasm after hyperosmotic stress: a novel Fps1-dependent response. Mol Biol Cell 14:3929–3941

Hieronymus H, Silver PA (2003) Genome-wide analysis of RNA-protein interactions illustrates specificity of the mRNA export machinery. Nat Genet 33:155–161

Hilleren P, McCarthy T, Rosbash M, Parker R, Jensen TH (2001) Quality control of mRNA 3′-end processing is linked to the nuclear exosome. Nature 413:538–542

Hobeika M, Brockmann C, Iglesias N, Gwizdek C, Neuhaus D, Stutz F et al (2007) Coordination of Hpr1 and ubiquitin binding by the UBA domain of the mRNA export factor *Mex67*. Mol Biol Cell 18:2561–2568

Hobeika M, Brockmann C, Gruessing F, Neuhaus D, Divita G, Stewart M et al (2009) Structural requirements for the ubiquitin-associated domain of the mRNA export factor *Mex67* to bind its specific targets, the transcription elongation THO complex component Hpr1 and nucleoporin FXFG repeats. J Biol Chem 284:17575–17583

Hodge CA, Colot HV, Stafford P, Cole CN (1999) Rat8p/Dbp5p is a shuttling transport factor that interacts with Rat7p/Nup159p and Gle1p and suppresses the mRNA export defect of xpo1-1 cells. EMBO J 18:5778–5788

Hodge CA, Tran EJ, Noble KN, Alcazar-Roman AR, Ben-Yishay R, Scarcelli JJ et al (2011) The Dbp5 cycle at the nuclear pore complex during mRNA export I: dbp5 mutants with defects in RNA binding and ATP hydrolysis define key steps for Nup159 and Gle1. Genes Dev 25:1052–1064

Houseley J, Tollervey D (2009) The many pathways of RNA degradation. Cell 136:763–776

Hsin JP, Manley JL (2012) The RNA polymerase II CTD coordinates transcription and RNA processing. Genes Dev 26:2119–2137

Huertas P, Aguilera A (2003) Co-transcriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. Mol Cell 12:711–721

Hurt E, Strasser K, Segref A, Bailer S, Schlaich N, Presutti C et al (2000) *Mex67p* mediates nuclear export of a variety of RNA polymerase II transcripts. J Biol Chem 275:8361–8368

Hurt E, Luo MJ, Rother S, Reed R, Strasser K (2004) Co-transcriptional recruitment of the serine-arginine-rich (SR)-like proteins Gbp2 and Hrb1 to nascent mRNA via the TREX complex. Proc Natl Acad Sci USA 101:1858–1862

Iglesias N, Tutucci E, Gwizdek C, Vinciguerra P, Von Dach E, Corbett AH et al (2010) Ubiquitin-mediated mRNP dynamics and surveillance prior to budding yeast mRNA export. Genes Dev 24:1927–1938

Inoue K, Mizuno T, Wada K, Hagiwara M (2000) Novel RING finger proteins, Air1p and Air2p, interact with Hmt1p and inhibit the arginine methylation of Npl3p. J Biol Chem 275:32793–32799

Jani D, Lutz S, Marshall NJ, Fischer T, Kohler A, Ellisdon AM et al (2009) Sus1, Cdc31, and the Sac3 CID region form a conserved interaction platform that promotes nuclear pore association and mRNA export. Mol Cell 33:727–737

Jensen TH, Boulay J, Rosbash M, Libri D (2001a) The DECD box putative ATPase Sub2p is an early mRNA export factor. Curr Biol 11:1711–1715

Jensen TH, Boulay J, Rosbash M, Libri D (2001b) The DECD box putative ATPase Sub2p is an early mRNA export factor. Curr Biol 11:1711–1715

Jensen TH, Dower K, Libri D, Rosbash M (2003) Early formation of mRNP: license for export or quality control? Mol Cell 11:1129–1138

Jia H, Wang X, Liu F, Guenther UP, Srinivasan S, Anderson JT et al (2011) The RNA helicase Mtr4p modulates polyadenylation in the TRAMP complex. Cell 145:890–901

Jimeno S, Aguilera A (2010) The THO complex as a key mRNP biogenesis factor in development and cell differentiation. J Biol 9:6

Jimeno S, Rondon AG, Luna R, Aguilera A (2002) The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability. EMBO J 21:3526–3535

Johnson SA, Cubberley G, Bentley DL (2009) Co-transcriptional recruitment of the mRNA export factor Yra1 by direct interaction with the 3′ end processing factor Pcf11. Mol Cell 33:215–226

Johnson SA, Kim H, Erickson B, Bentley DL (2011) The export factor Yra1 modulates mRNA 3′ end processing. Nat Struct Mol Biol 18:1164–1171

Kadowaki T, Chen S, Hitomi M, Jacobs E, Kumagai C, Liang S et al (1994) Isolation and characterization of *Saccharomyces cerevisiae* mRNA transport-defective (mtr) mutants [published erratum appears in J Cell Biol 1994 Sep; 126(6):1627]. J Cell Biol 126:649–659

Katahira J (2012) mRNA export and the TREX complex. Biochim Biophys Acta 1819:507–513

Kim Guisbert K, Duncan K, Li H, Guthrie C (2005) Functional specificity of shuttling hnRNPs revealed by genome-wide analysis of their RNA binding profiles. RNA 11:383–393

Kim M, Ahn SH, Krogan NJ, Greenblatt JF, Buratowski S (2004) Transitions in RNA polymerase II elongation complexes at the 3′ ends of genes. EMBO J 23:354–364

Kohler A, Hurt E (2007) Exporting RNA from the nucleus to the cytoplasm. Nat Rev Mol Cell Biol 8:761–773

Kong KYE, Tang HMV, Pan K, Huang Z, Lee THJ, Hinnebusch AG et al (2013) Co-transcriptional recruitment of yeast TRAMP complex to intronic sequences promotes optimal pre-mRNA splicing. Nucleic Acids Res 42(1):643–660

Krebber H, Taura T, Lee MS, Silver PA (1999) Uncoupling of the hnRNP Npl3p from mRNAs during the stress-induced block in mRNA export. Genes Dev 13:1994–2004

Kress TL, Krogan NJ, Guthrie C (2008) A single SR-like protein, Npl3, promotes pre-mRNA splicing in budding yeast. Mol Cell 32:727–734

Lee MS, Henry M, Silver PA (1996) A protein that shuttles between the nucleus and the cytoplasm is an important mediator of RNA export. Genes Dev 10:1233–1246

Lei EP, Silver PA (2002) Intron status and 3′-end formation control co-transcriptional export of mRNA. Genes Dev 16:2761–2766

Lei EP, Krebber H, Silver PA (2001) Messenger RNAs are recruited for nuclear export during transcription. Genes Dev 15:1771–1782

Lei EP, Stern CA, Fahrenkrog B, Krebber H, Moy TI, Aebi U et al (2003) Sac3 is an mRNA export factor that localizes to cytoplasmic fibrils of nuclear pore complex. Mol Biol Cell 14:836–847

Lewis JD, Izaurralde E (1997) The role of the cap structure in RNA processing and nuclear export. Eur J Biochem 247:461–469

Lewis A, Felberbaum R, Hochstrasser M (2007) A nuclear envelope protein linking nuclear pore basket assembly, SUMO protease regulation, and mRNA surveillance. J Cell Biol 178:813–827

Lidschreiber M, Leike K, Cramer P (2013) Cap completion and C-terminal repeat domain kinase recruitment underlie the initiation-elongation transition of RNA polymerase II. Mol Cell Biol 33:3805–3816

Luna R, Jimeno S, Marin M, Huertas P, Garcia-Rubio M, Aguilera A (2005) Interdependence between transcription and mRNP processing and export, and its impact on genetic stability. Mol Cell 18:711–722

Luthra R, Kerr SC, Harreman MT, Apponi LH, Fasken MB, Ramineni S et al (2007) Actively transcribed GAL genes can be physically linked to the nuclear pore by the SAGA chromatin modifying complex. J Biol Chem 282:3042–3049

Ma WK, Cloutier SC, Tran EJ (2013) The DEAD-box protein Dbp2 functions with the RNA-binding protein Yra1 to promote mRNP assembly. J Mol Biol 425:3824–3838

MacKellar AL, Greenleaf AL (2011) Co-transcriptional association of mRNA export factor Yra1 with C-terminal domain of RNA polymerase II. J Biol Chem 286:36385–36395

McBride AE, Cook JT, Stemmler EA, Rutledge KL, McGrath KA, Rubens JA (2005) Arginine methylation of yeast mRNA-binding protein Npl3 directly affects its function, nuclear export, and intranuclear protein interactions. J Biol Chem 280:30888–30898

Meyer M, Vilardell J (2009) The quest for a message: budding yeast, a model organism to study the control of pre-mRNA splicing. Brief Funct Genomic Proteomic 8:60–67

Morris DP, Greenleaf AL (2000) The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. J Biol Chem 275:39935–39943

Murphy R, Watkins JL, Wente SR (1996) GLE2, a Saccharomyces cerevisiae homologue of the Schizosaccharomyces pombe export factor RAE1, is required for nuclear pore complex structure and function. Mol Biol Cell 7:1921–1937

Neville M, Rosbash M (1999) The NES-Crm1p export pathway is not a major mRNA export route in *Saccharomyces cerevisiae*. EMBO J 18:3746–3756

Niepel M, Molloy KR, Williams R, Farr JC, Meinema AC, Vecchietti N et al (2013) The nuclear basket proteins Mlp1p and Mlp2p are part of a dynamic interactome including Esc1p and the proteasome. Mol Biol Cell 24(24):3920–3938

Noble KN, Tran EJ, Alcazar-Roman AR, Hodge CA, Cole CN, Wente SR (2011) The Dbp5 cycle at the nuclear pore complex during mRNA export II: nucleotide cycling and mRNP remodeling by Dbp5 are controlled by Nup159 and Gle1. Genes Dev 25:1065–1077

Ohno M, Segref A, Bachi A, Wilm M, Mattaj IW (2000) PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. Cell 101:187–198

Palancade B, Zuccolo M, Loeillet S, Nicolas A, Doye V (2005) Pml39, a novel protein of the nuclear periphery required for nuclear retention of improper messenger ribonucleoparticles. Mol Biol Cell 16:5258–5268

Pante N, Aebi U (1995) Exploring nuclear pore complex structure and function in molecular detail. J Cell Sci Suppl 19:1–11

Pascual-Garcia P, Govind CK, Queralt E, Cuenca-Bono B, Llopis A, Chavez S et al (2008) Sus1 is recruited to coding regions and functions during transcription elongation in association with SAGA and TREX2. Genes Dev 22:2811–2822

Pena A, Gewartowski K, Mroczek S, Cuéllar J, Szykowska A, Prokop A et al (2012) Architecture and nucleic acids recognition mechanism of the THO complex, an mRNP assembly factor. EMBO J 31:1605–1616

Perez-Canadillas JM (2006) Grabbing the message: structural basis of mRNA 3′UTR recognition by Hrp1. EMBO J 25:3167–3178

Porrua O, Libri D (2013) RNA quality control in the nucleus: the Angels' share of RNA. Biochim Biophys Acta 1829:604–611

Pritchard CE, Fornerod M, Kasper LH, van Deursen JM (1999) RAE1 is a shuttling mRNA export factor that binds to a GLEBS-like NUP98 motif at the nuclear pore complex through multiple domains. J Cell Biol 145:237–254

Qu X, Lykke-Andersen S, Nasser T, Saguez C, Bertrand E, Jensen TH et al (2009) Assembly of an export-competent mRNP is needed for efficient release of the 3′-end processing complex after polyadenylation. Mol Cell Biol 29:5327–5338

Rodriguez-Navarro S, Hurt E (2011) Linking gene regulation to mRNA production and export. Curr Opin Cell Biol 23:302–309

Rodriguez-Navarro S, Strasser K, Hurt E (2002) An intron in the YRA1 gene is required to control Yra1 protein expression and mRNA export in yeast. EMBO Rep 3:438–442

Rodriguez-Navarro S, Fischer T, Luo MJ, Antunez O, Brettschneider S, Lechner J et al (2004) Sus1, a functional component of the SAGA histone acetylase complex and the nuclear pore-associated mRNA export machinery. Cell 116:75–86

Rollenhagen C, Hodge CA, Cole CN (2007) Following temperature stress, export of heat shock mRNA occurs efficiently in cells with mutations in genes normally important for mRNA export. Eukaryot Cell 6(3):505–513

Rondon AG, Jimeno S, Garcia-Rubio M, Aguilera A (2003) Molecular evidence that the eukaryotic THO/TREX complex is required for efficient transcription elongation. J Biol Chem 278:39037–39043

Rougemaille M, Dieppois G, Kisseleva-Romanova E, Gudipati RK, Lemoine S, Blugeon C et al (2008) THO/Sub2p functions to coordinate 3′-end processing with gene-nuclear pore association. Cell 135:308–321

Rout MP, Aitchison JD, Suprapto A, Hjertaas K, Zhao Y, Chait BT (2000) The yeast nuclear pore complex: composition, architecture, and transport mechanism. J Cell Biol 148:635–651

Saavedra C, Tung KS, Amberg DC, Hopper AK, Cole CN (1996) Regulation of mRNA export in response to stress in *Saccharomyces cerevisiae*. Genes Dev 10:1608–1620

Saavedra CA, Hammell CM, Heath CV, Cole CN (1997) Yeast heat shock mRNAs are exported through a distinct pathway defined by Rip1p. Genes Dev 11:2845–2856
Saguez C, Schmid M, Olesen JR, Ghazy MA, Qu X, Poulsen MB et al (2008) Nuclear mRNA surveillance in THO/sub2 mutants is triggered by inefficient polyadenylation. Mol Cell 31:91–103
Sampathkumar P, Kim SJ, Manglicmot D, Bain KT, Gilmore J, Gheyi T et al (2012) Atomic structure of the nuclear pore complex targeting domain of a Nup116 homologue from the yeast, Candida glabrata. Proteins 80:2110–2116
San Paolo S, Vanacova S, Schenk L, Scherrer T, Blank D, Keller W et al (2009) Distinct roles of non-canonical Poly(A) polymerases in RNA metabolism. PLoS Genet 5:e1000555
Shen EC, Henry MF, Weiss VH, Valentini SR, Silver PA, Lee MS (1998) Arginine methylation facilitates the nuclear export of hnRNP proteins. Genes Dev 12:679–691
Skruzny M, Schneider C, Racz A, Weng J, Tollervey D, Hurt E (2009) An endoribonuclease functionally linked to perinuclear mRNP quality control associates with the nuclear pore complexes. PLoS Biol 7:e8
Snay-Hodge CA, Colot HV, Goldstein AL, Cole CN (1998) Dbp5p/Rat8p is a yeast nuclear pore-associated DEAD-box protein essential for RNA export. EMBO J 17:2663–2676
Soucek S, Corbett AH, Fasken MB (2012) The long and the short of it: the role of the zinc finger polyadenosine RNA binding protein, Nab2, in control of poly(A) tail length. Biochim Biophys Acta 1819:546–554
Stage-Zimmermann T, Schmidt U, Silver PA (2000) Factors affecting nuclear export of the 60S ribosomal subunit in vivo. Mol Biol Cell 11:3777–3789
Stewart M (2010) Nuclear export of mRNA. Trends Biochem Sci 35:609–617
Strambio-De-Castillia C, Niepel M, Rout MP (2010a) The nuclear pore complex: bridging nuclear transport and gene regulation. Nat Rev Mol Cell Biol 11:490–501
Strambio-De-Castillia C, Niepel M, Rout MP (2010b) The nuclear pore complex: bridging nuclear transport and gene regulation. Nat Rev Mol Cell Biol 11:490–501
Strasser K, Hurt E (2000) Yra1p, a conserved nuclear RNA-binding protein, interacts directly with *Mex67p* and is required for mRNA export. EMBO J 19:410–420
Strasser K, Hurt E (2001) Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p. Nature 413:648–652
Strasser K, Bassler J, Hurt E (2000) Binding of the *Mex67p/Mtr2p* heterodimer to FXFG, GLFG, and FG repeat nucleoporins is essential for nuclear mRNA export. J Cell Biol 150:695–706
Strasser K, Masuda S, Mason P, Pfannstiel J, Oppizzi M, Rodriguez-Navarro S et al (2002) TREX is a conserved complex coupling transcription with messenger RNA export. Nature 417:304–308
Tarn WY, Hsu CH, Huang KT, Chen HR, Kao HY, Lee KR et al (1994) Functional association of essential splicing factor(s) with PRP19 in a protein complex. EMBO J 15:2421–2431
Tieg B, Krebber H (2013) Dbp5—from nuclear export to translation. Biochim Biophys Acta 1829:791–798
Tran EJ, Zhou Y, Corbett AH, Wente SR (2007) The DEAD-box protein Dbp5 controls mRNA export by triggering specific RNA: protein remodeling events. Mol Cell 28:850–859
Tseng SS, Weaver PL, Liu Y, Hitomi M, Tartakoff AM, Chang TH (1998) Dbp5p, a cytosolic RNA helicase, is required for poly(A)+ RNA export. EMBO J 17:2651–2662
Tuck AC, Tollervey D (2013) A transcriptome-wide atlas of RNP composition reveals diverse classes of mRNAs and lncRNAs. Cell 154:996–1009
Tutucci E, Stutz F (2011) Keeping mRNPs in check during assembly and nuclear export. Nat Rev Mol Cell Biol 12:377–384
Wahl MC, Will CL, Luhrmann R (2009) The spliceosome: design principles of a dynamic RNP machine. Cell 136:701–718
Weirich CS, Erzberger JP, Flick JS, Berger JM, Thorner J, Weis K (2006) Activation of the DExD/H-box protein Dbp5 by the nuclear-pore protein Gle1 and its coactivator InsP6 is required for mRNA export. Nat Cell Biol 8:668–676

Wente SR, Blobel G (1993) A temperature-sensitive NUP116 null mutant forms a nuclear envelope seal over the yeast nuclear pore complex thereby blocking nucleocytoplasmic traffic. J Cell Biol 123:275–284

Will CL, Luhrmann R (2011) Spliceosome structure and function. Cold Spring Harb Perspect Biol 3:a003707

Windgassen M, Krebber H (2003) Identification of Gbp2 as a novel poly(A)+ RNA-binding protein involved in the cytoplasmic delivery of messenger RNAs in yeast. EMBO Rep 4:278–283

Windgassen M, Sturm D, Cajigas IJ, Gonzalez CI, Seedorf M, Bastians H et al (2004) Yeast shuttling SR proteins Npl3p, Gbp2p, and Hrb1p are part of the translating mRNPs, and Npl3p can function as a translational repressor. Mol Cell Biol 24:10479–10491

Wong CM, Qiu H, Hu C, Dong J, Hinnebusch AG (2007) Yeast cap binding complex impedes recruitment of cleavage factor IA to weak termination sites. Mol Cell Biol 27:6520–6531

Yao W, Roser D, Kohler A, Bradatsch B, Bassler J, Hurt E (2007) Nuclear export of ribosomal 60S subunits by the general mRNA export receptor *Mex67-Mtr2*. Mol Cell 26:51–62

Zenklusen D, Vinciguerra P, Strahm Y, Stutz F (2001) The yeast hnRNP-Like proteins Yra1p and Yra2p participate in mRNA export through interaction with *Mex67p*. Mol Cell Biol 21:4219–4232

Zheng C, Fasken MB, Marshall NJ, Brockmann C, Rubinson ME, Wente SR et al (2010) Structural basis for the function of the *Saccharomyces cerevisiae* Gfd1 protein in mRNA nuclear export. J Biol Chem 285:20704–20715

Chapter 5
mRNA Translation: Fungal Variations on a Eukaryotic Theme

Tobias von der Haar and Leoš Shivaya Valášek

Abstract The accurate transfer of information from a nucleotide-based code to a protein-based one is at the heart of all life processes. The actual information transfer occurs during protein synthesis or translation, and is catalysed by ribosomes, supported by a large host of additional protein activities—the translation factors. This chapter reviews how the different eukaryotic initiation, elongation and termination factors assist the ribosome in establishing appropriate contacts with mRNAs during translation initiation, decode the genetic code during translation elongation, and finally release the newly made polypeptide and reuse the ribosomes during the termination and recycling phases.

Contents

T. von der Haar (✉)
Kent Fungal Group, School of Biosciences, University of Kent, Canterbury CT2 7NJ, UK
e-mail: T.von-der-Haar@kent.ac.uk

L. S. Valášek
Laboratory of Regulation of Gene Expression, Institute of Microbiology ASCR, Videnska 1083 142 20 Prague, The Czech Republic
e-mail: valasekl@biomed.cas.cz

A. Sesma and T. von der Haar (eds.), *Fungal RNA Biology*,
DOI: 10.1007/978-3-319-05687-6_5,

Introduction

Like the genetic code itself, the mechanism of codon decoding and protein synthesis relies on universally conserved factors and biochemical reactions, with specific variations that have been introduced in different evolutionary groups of organisms. Of the three phases of translation (initiation, elongation and termination), elongation is the most highly conserved and uses components and reactions that are recognisable from prokaryotes to eukaryotes. In contrast, initiation and termination differ significantly between eukaryotes and prokaryotes. Fungi generally rely on the canonical eukaryotic pathways, albeit with some specific alterations.

The series of molecular reactions that make up translation is described here in sufficient detail to provide a broad overview, although it is impossible to describe this topic in a single chapter with any molecular depth. Our understanding of translation has made enormous progress in the last few years, thanks in no small part to studies on the mechanism of translation in baker's yeast. With the recent rapid progress, the individual steps of translation have been subject to excellent in depth reviews, which are cited in each section for further reading.

The central player in translation in all organisms is the ribosome, the largest non-membrane bound molecular structure in a cell (Fig. 5.1). The ribosome consists of two subunits: in eukaryotes, these are termed the small or 40S subunit, which consists of the 18S ribosomal RNA (rRNA) and around 30 ribosomal proteins, and the 60S subunit, which consists of 3 rRNA molecules (the 5S, 5.8S and 28S rRNA) and 40–50 ribosomal proteins. The assembled 80S ribosome contains an mRNA channel and three tRNA binding sites termed the A-, P- and E-site (so named because at the beginning of the decoding process they contain the aminoacyl-tRNA, peptidyl-tRNA, or are empty).

The Molecular Mechanism of Translation

Translation Initiation

Translation initiation starts with the assembly of two large ribonucleoprotein complexes on the small ribosomal subunit and the mRNA, respectively (Fig. 5.2). The main players in these complexes are the eukaryotic initiation factors (eIFs).

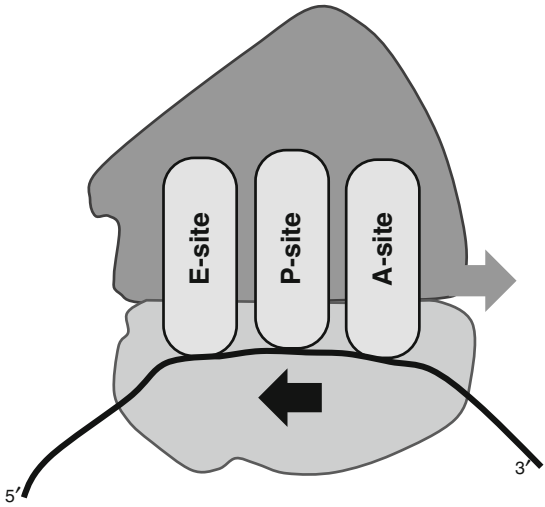
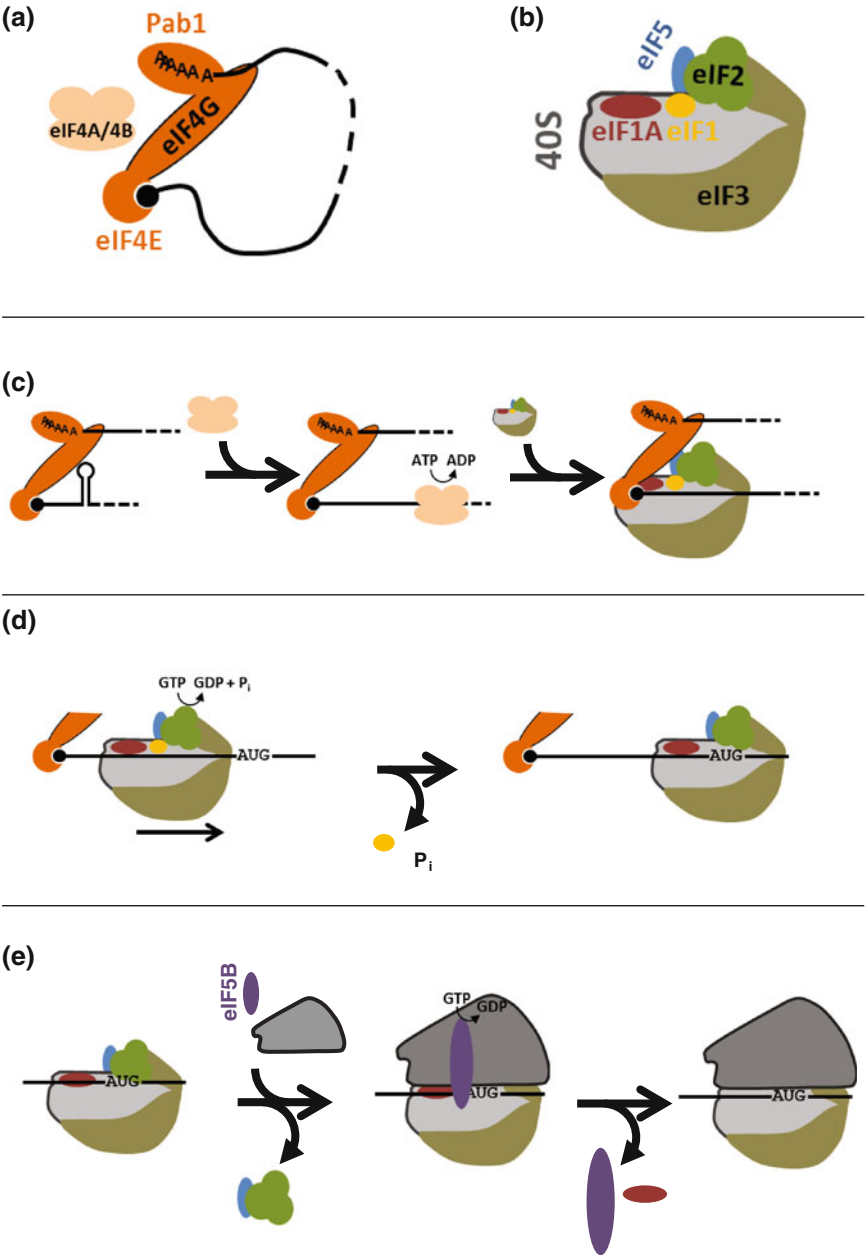


Fig. 5.1 The schematic structure of a ribosome. Shown are the small or 40S subunit in *light grey*, the large or 60S subunit in *dark grey*, and the mRNA in *black*. The three tRNA binding sites are outlined, and the direction of movement of the mRNA (*black arrow*) versus the ribosome (*grey arrow*) is indicated

The mRNA-bound complex is commonly referred to as the cap-binding or eIF4F complex, and comprises the cap-binding protein eIF4E and the adaptor protein eIF4G as core components (Topisirovic et al. 2011) (Fig. 5.2a). Other proteins interact with the core complex, including the poly(A) binding protein, Pab1. eIF4E and Pab1 bind molecular features located at the extreme ends of the message, namely the 5′ m7G cap structure and the 3′ poly(A) tail. Through the bridging function of eIF4G which binds both eIF4E and Pab1, as well as multiple direct mRNA contacts of eIF4G, mRNAs can be effectively circularised, although it is currently under debate to what extent this occurs in vivo. The cap-binding complex recruits RNA helicases that interact more weakly with eIF4G (Korneeva et al. 2001; He et al. 2003) and therefore cycle in and out of the complex (Pause et al. 1994). The central helicase activity is provided by the archetypal RNA helicase eIF4A, and is stimulated by interaction with eIF4G as well as another co-factor, eIF4B (a second mammalian co-factor, eIF4H, has no detectable homologues in most fungi). The helicase activity associated with these factors is thought to remove secondary structure immediately adjacent to the cap structure, thus clearing a landing zone that is free from obstruction for the initiating ribosomes.

The second complex forms through the mutually stimulated association of multiple translation factors, eIF1, eIF1A, eIF2, eIF3 and eIF5, with the small ribosomal subunit (Valásek 2012; Aitken and Lorsch 2012) (Fig. 5.2b). The resulting complex is referred to as the 43S pre-initiation complex (PIC). eIF2 and eIF3 are themselves multimeric protein complexes, and in most fungi the full mass of initiation factors interacting with the 1.2GDa small ribosomal subunit is around 500 kDa (in mammals the complex is larger due to a higher number of eIF3 subunits). Key molecular functions for the 43S PIC include ensuring proper placement of the initiator methionyl-$tRNA_i^{Met}$ in the ribosomal P-site, a pre-requisite for sensitive detection of the start codon, and opening of the mRNA


(a)
Pab1
eIF4A/4B
eIF4G
eIF4E
(b)
eIF5
eIF2
40S
eIF1A
eIF1
eIF3
(c)
ATP ADP
(d)
GTP GDP + Pi
AUG
Pi
AUG
(e)
eIF5B
AUG
GTP
GDP
AUG
AUG

◀ **Fig. 5.2** The canonical translation initiation pathway. **a** The schematic structure of the cap-binding complex. The translation initiation factors eIF4E and eIF4G, and the poly(A) binding protein Pab1, form a complex that connects the ends of mRNAs. eIF4A and eIF4B, which provide RNA helicase activity to the complex, interact transiently with the core cap complex components. **b** illustration of the ribosome-associated initiation factor complex. eIF2 is present in this complex as a ternary complex with the initiator tRNA and GTP. The majority of factors is bound on the interface-site of the 40S subunit around the tRNA located in the P-site. In contrast, eIF3 is bound on the solvent side of the ribosome, with extensions that contact the interface-side bound factors. **c** Removal of secondary structure from the mRNA 5′-end by helicases allows the cap-binding complex to recruit the 40S: initiation factor complex. **d** During scanning, the ribosome moves along the 5′-UTR until it encounters a start codon in a favourable sequence context. This leads to the release of eIF1 and a phosphate produced by GTP hydrolysis on eIF2, and the arrest of the scanning complex. **e** Following dissociation of eIF5 and eIF2, the large ribosomal subunit is recruited with the help of eIF1A and eIF5B. These factors then also dissociate, leaving an elongation-competent, fully assembled ribosome placed over the start codon. eIF3 is thought to remain associated with the ribosome at least for the initial elongation cycles

binding channel in the 40S subunit, a prerequisite for recruitment of eIF4F-bound mRNAs to the small subunit.

Contact between the 43S PIC and eIF4F-bound mRNA is initiated by contacts between their associated proteins (Fig. 5.2c) (Valášek 2012). The details of these initial contacts may differ between fungi and mammals, since the eIF4G:eIF3 interaction, thought to be responsible for 40S recruitment in mammals, could not be detected for the homologous baker's yeast proteins, where eIF3 by itself seems to play the most critical role (Jivotovskaya et al. 2006). An interaction between eIF4B and eIF3 was proposed to further reinforce contacts during the initial PIC assembly (Méthot et al. 1996). In agreement, yeast eIF4B was recently shown to share one of the eIF3 binding sites on the solvent-exposed side of the 40S subunit, and to promote productive interactions between eIF4A and the 43S·mRNA PIC and efficient mRNA recruitment (Walker et al. 2013).

Following formation of the 48S PIC, the resulting complex adopts a so-called scanning conducive conformation. This enables the PIC to move along the 5′-untranslated region (5′-UTR) in a process known as scanning. Scanning *per se* consists of unwinding of any secondary structures that may impede ribosome movement, and inspection of the mRNA's 5′ leader in search for AUG start codons. The exact mode of movement is poorly characterised, but most likely involves energy-dependent assistance by DNA helicases other than—or in addition to—eIF4A, such as the fungal Ded1 (Berthelot et al. 2004) and mammalian DHX29 (Pisareva et al. 2008). Alternative models posit that the scanning subunit may be given directionality by a ratchet-like action of eIF4B, which follows the scanning ribosome and prevents backwards movement (Spirin 2009).

In the scanning conducive conformation, the anticodon of the initiator $tRNA_i^{Met}$ is not fully engaged in the ribosomal P-site. This is thought to prevent premature engagement with near-cognate start codons. eIF2 partially hydrolyzes its GTP with the help of eIF5; however, prior to start codon recognition, the "gate-keeping" function of eIF1 prevents the release of the resulting phosphate ion. The presence of an AUG start codon in the P-site then triggers displacement of eIF1 from the

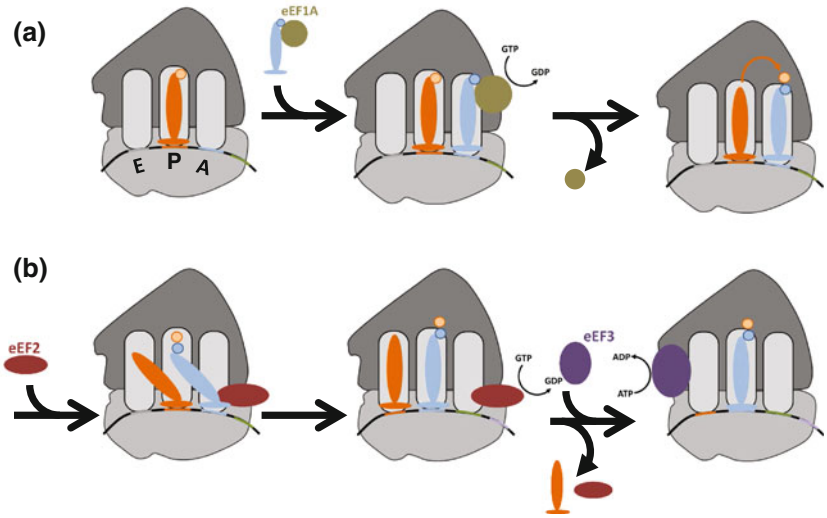


Fig. 5.3 The principal events of translation elongation. **a** During the decoding and peptidyl transfer stage of elongation, tRNA:eEF1A complexes enter the ribosomal A-site. Non-cognate complexes dissociate again, whereas cognate complexes lead to formation of a stable ribosome:tRNA complex and GTP hydrolysis on eEF1A. Upon dissociation of eEF1A, the ribosome catalyses peptidyl transfer of the P-site amino acid onto the A-site tRNA. **b** During the translocation stage of elongation, eEF2 then binds to the ribosome, leading to partial translocation of the tRNAs which adopt the so-called hybrid state. GTP hydrolysis on eEF2 induces complete translocation, followed by eEF3-catalysed E-site tRNA release. The cycle then recommences with the sampling of tRNA:eEF1A complexes in the A-site

PIC, followed by release of free Pi and full accommodation of the initiator Met-tRNA$_i^{Met}$ in the P-site (Saini et al. 2010) (Fig. 5.2d). Together these changes terminate the scanning movement and arrest the small subunit with the start codon located in its P-site. The events from formation of the small subunit complex to recognition of the start codon have been review in depth in (Valášek 2012; Aitken and Lorsch 2012).

Upon AUG recognition, all initiation factors except eIF1A and eIF3 leave the PIC (at this point, the PIC may also still be in contact with the cap-binding complex (Szamecz et al. 2008)). An additional factor, eIF5B, then binds the small subunit and catalyses recruitment of the large subunit (Fig. 5.2e). GTP hydrolysis on eIF5B is required for the 80S ribosome to adopt the final, elongation-competent conformation, and this is aided by contacts with eIF1A (Acker et al. 2009).

Several of the translation factors acting in the initiation pathway have ATPase or GTPase activity, and following an initiation event, the resulting factor: NDP complexes must be recycled to their NTP-bound form to allow subsequent initiation events to occur. Both for eIF4A and eIF5B, this nucleotide exchange is thought to occur passively through the natural excess of triphosphate nucleotides over diphosphate nucleotides in the cell. In contrast, regeneration of GTP-bound eIF2 involves a sophisticated guanidine exchange cycle. At the core of this cycle is

a dedicated guanidine nucleotide exchange factor (GEF) for eIF2, the five-subunit factor eIF2B, which upon binding to eIF2 reduces this factor's affinity for GDP (Mohammed-Qureshi et al. 2008). eIF2B activity itself is regulated by interactions between eIF5 and the GDP-bound eIF2 (Jennings and Pavitt 2010).

Translation Elongation

The elongation-competent ribosome emerging from the initiation pathway undergoes a cyclical series of reactions that lead to decoding of the A-site codon, transfer of the amino acid or peptide bound to the P-site tRNA onto the amino acid bound to the A-site tRNA, and translocation of the ribosome onto the following codon (Fig. 5.3).

Codon decoding occurs through the action of tRNAs, which are delivered to the ribosomal A-site in a ternary complex with the elongation factor *eEF1A* and GTP. Most organisms contain 40–50 functional tRNA species (i.e., tRNAs with distinct anticodons). The identity of the incoming tRNA is usually random, although it has been observed that in some cases the ribosome can bias the nature of the incoming tRNA and increase the probability that a species matching the codon in the A-site enters the ribosome (Cannarrozzi et al. 2010).

The incoming tRNA can belong to any one of the following three groups. Which group this is determines the fate of the ternary complex in the ribosome. *Non-cognate* tRNAs have an anticodon that is not compatible to form extensive base pairs with the A-site codon, and these tRNAs rapidly leave the ribosome again. Similarly, *near-cognate* tRNAs are not normally competent to decode the A-site codon, but their base-pairing properties are such that they can undergo part of the reactions normally reserved for cognate tRNAs, and near-cognate tRNAs therefore typically occupy the A-site for longer before they dissociate again. The last group, *cognate* tRNAs, have an anticodon that forms Watson-Crick basepairs with the first and second base of the A-site codon, and forms either Watson-Crick or so-called wobble-base pairs with the third base of the codon. Wobble-base pairing is often facilitated by the posttranscriptional modification of the tRNA to contain unusual bases like inosine or mcm^5s^2-uridine (Agris et al. 2007). These modifications and the resulting ability to wobble-decode can allow one tRNA species to decode multiple codons that end in different bases, enabling the decoding of 61 sense codons with only 40–50 tRNA species.

In the ribosomal A-site, ternary complexes of cognate tRNAs undergo a series of reactions that produce structural changes in the A-site, induction of GTP hydrolysis by eEF1A, exit of eEF1A from the ribosome in its GDP-bound form, and finally transfer of the peptide from the P-site tRNA onto the amino acid bound to the A-site tRNA by the peptidyl transferase centre (Fig. 5.3). The net result of these reactions is the extension of the nascent polypeptide by one amino acid.

The molecular details of codon decoding are covered in several recent reviews (Rodnina and Wintermeyer 2009; Vorhees and Ramakrishnan 2013).

For some sequences, specifically multiple consecutive proline codons, peptidyl transfer is much slower than for other amino acids and this can lead to ribosome stalling. In these cases, the assistance of an additional factor, termed eIF5A, is required to facilitate the peptidyl transfer reaction (eIF5A was initially characterised as an initiation factor, hence the designation as eIF rather than eEF). Structural modelling suggested that an unusual, posttranslationally modified hypusine residue in eIF5A contacts the peptidyl tRNA acceptor stem, and this may thus slightly alter either the conformation of this tRNA or the energetics of the tRNA:peptide bond to help peptidyl transfer under unfavourable conditions (Gutierrez et al. 2013).

Translocation then requires assistance from another elongation factor, eEF2. Binding of this factor in complex with GTP leads to a conformational change in the tRNAs which adopt the so-called hybrid state, with the acceptor stems of the P- and A-site tRNAs now located in the E- and P-sites, respectively, and this forms the first part of the translocation reaction. Hydrolysis of the eEF2-bound GTP then leads to a conformational change on this factor that pushes the anticodon ends of the tRNAs backwards in the ribosome, so that the empty tRNA is now fully located in the E-site, and the peptide-bound tRNA is fully located in the P-site. In fungi, the following release of the E-site tRNA is catalysed by a non-canonical and fungal-specific elongation factor eEF3, in an ATPase-dependent manner (Andersen et al. 2006) (in non-fungal eukaryotes, it was suggested that an ATPase activity resident in the ribosome itself may provide a similar functionality (Miyazaki and Kagiyama 1990) although this has later been questioned (Kovalchuke et al. 1995)). The molecular details of the translocation reaction are again covered in recent reviews (Vorhees and Ramakrishnan 2013).

Although release of the E-site tRNA completes the decoding cycle from the point of view of the ribosome, several activities need to be regenerated in order to sustain subsequent translational activity. The GDP/ADP bound forms of eEF2 and eEF3 are thought to be recycled to the GTP/ATP-bound forms without involvement of a nucleotide exchange factor. In contrast, guanidine exchange on eEF1A is catalysed by a dedicated GEF, the multi-subunit eEF1B. In yeast, this factor has a simpler subunit structure than in higher eukaryotes, consisting of one regulatory subunit (eEF1Bγ) and one catalytic subunit (eEF1Bα) (Le Sourd et al. 2006).

Lastly, the deacylated tRNA must be regenerated into its aminoacylated form, which is catalysed by the aminoacyl-tRNA synthetases (Ling et al. 2009). Fungi, like most eukaryotes, contain one synthetase for each amino acid, which catalyses aminoacylation of all tRNA species decoding a group of synonymous codons. The ability of one synthetase to recognise all its cognate tRNAs but to reject all other species is mediated by a complex system of positive and negative sequence elements which are distributed throughout the tRNA body (Giegé et al. 1998).

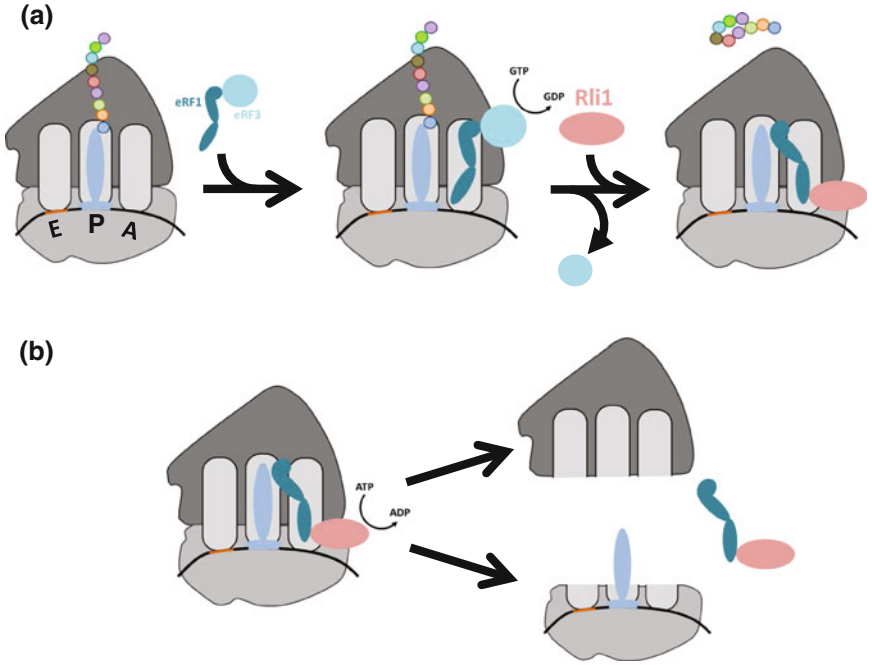


Fig. 5.4 Translation termination and ribosome recycling. **a** During translation termination stop codons entering the ribosomal A-site are decoded by a protein complex consisting of the release factors eRF1 and eRF3. eRF1 is thought to act analogously to the tRNA, and eRF3 analogously to eEF1A, during sense codon decoding. If the A-site codon is a stop codon, the eRF3-bound GTP is hydrolysed to GDP. eRF3 is then exchanged for Rli1 in a reaction likely assisted by Hcr1, which activates eRF1 to induce severing of the tRNA:peptide bond by the ribosomal peptidyl transferase centre. **b** During ribosome recycling, ATP hydrolysis on Rli1 generates a power stroke that dissociates the ribosome into its two subunits. The remaining small subunit:tRNA:mRNA complex is then dissociated through the action of initiation factors which prepare the subunit for the next round of translation

Translation Termination and Ribosome Recycling

The translation elongation cycle is repeated until one of the three stop codons (UAA, UGA or UAG) enters the ribosomal A-site (von der Haar and Tuite 2007; Jackson et al. 2012). There are normally no cognate tRNAs available to decode these codons (exceptions are the rare cases of suppressor tRNAs, which can arise from mutations in the anticodon of normal tRNAs). Instead, stop codons are decoded by a proteinaceous decoding factor termed release factor 1 or eRF1 (Fig. 5.4). During codon decoding, complexes between eRF1 and its binding partner, eRF3, are thought to be processed in a manner that is similar to eEF1A:tRNA:GTP ternary complexes, i.e., they enter the ribosomal A-site but are released again if the A-site codon is not a stop codon.

If the A-site codon is a stop codon, the eRF1:eRF3 complex is tightly bound by the ribosome, and recognition of the stop codon by the N-terminal domain of eRF1 (the functional equivalent of the tRNA anticodon) is communicated to eRF3. eRF3 then hydrolyses its bound GTP and is ejected from the ribosome:eRF1 complex, in a reaction stimulated by Hcr1. While this eIF3-related factor stimulates translation termination, the core-eIF3 complex antagonizes termination, likely by interacting with eRF1 and inhibiting its decoding activity (Beznoskova et al. 2013). eRF3 release allows binding of an ATP-bound four iron-sulphur cluster protein, Rli1, to eRF1, leading to a conformational change in the release factor which then essentially stimulates the ribosomal peptidyl transferase centre to transfer the nascent peptide onto a water molecule, thereby releasing it from the ribosome (Shoemaker and Green 2011).

The peptide-free complex of ribosome, mRNA and deacylated tRNA that results from the termination reaction is stable, and must be resolved in a series of further recycling steps. The first step of ribosome recycling is again promoted by Rli1. This factor hydrolyses its bound ATP, and translates the energy from this reaction into a power stroke that is thought to separate the ribosome into a separate 40S:tRNA:mRNA complex and a free 60S subunit (Becker et al. 2012). In a final reaction, eIFs 3, 1 and 1A and Hcr1 were proposed to recycle the 40S:tRNA:mRNA complex, at least in a mammalian in vitro system (Jackson et al. 2012).

Round and Round: Translation on Circularised Messages

Due to the action of the cap-binding complex which contacts both the 5′ cap structure and the 3′ poly(A) tail, mRNAs can be effectively circularised. This circularisation has been conclusively demonstrated in vitro (Wells et al. 1998), and can also be observed on mRNAs isolated from mammalian cells (Christensen et al. 1987). In cell free extracts, the dynamics of ribosome incorporation into polysomes indicate that ribosomes go around the same circularised message several times, without exchanging with the soluble pool of ribosomes (Kopeina et al. 2008). Several molecular contacts are known that might facilitate re-initiation, including contacts between the release factors and the poly(A) binding protein (Cosson et al. 2002; Hosoda et al. 2003; Amrani et al. 2008) and between the release factors and eIF3 (Beznoskova et al. 2013). These contacts may conceivably accelerate the formation of new pre-initiation complexes, thus either passively or actively favouring re-initiation in the vicinity of the termination site.

However, the details of this re-initiation mode of translation initiation, and how it differs from the de novo initiation mode of cytoplasmic ribosomes described above, are still poorly understood. Moreover, it is currently unclear whether all mRNAs or only a subset are circular in vivo, and to what extend the circularisation affects the efficiency of their expression. In yeast, the Pab1–eIF4G interaction is dispensable for wild-type cell growth (Park et al. 2011a). This suggested that either the key determinant of the closed-loop formation (at least in yeast) is the Pab1

interaction with other factor(s) than eIF4G, or the overall contribution of the closed-loop is only stimulatory and might become more important under certain cellular conditions such as stress.

Global Regulation of Translational Activity

The purpose of translation is to generate a proteome that maximises the chance of cells to survive and thrive in the different conditions they encounter. To achieve this, translation must be regulated at different levels. Global regulation of protein synthesis is necessary to match translation with other biosynthetic activities, thereby maintaining the average chemical composition of cells through the cell division cycle. Global regulation also matches the high energy demand of translation to available resources. On the other hand, mRNA specific regulation of translation controls the mixture of proteins that results from a given transcriptome. Global and mRNA specific modes of control are interlinked and cannot be fully separated, but for the purpose of clarity this section will focus on more global mechanisms, whereas the following section will focus on features of mRNAs that are used for their specific control.

Since the initiation factor activities described above are essential for on-going translation, they can be used to control global translational activity. Two major regulatory hubs among the translation initiation factors are the cap-binding protein eIF4E, and the trimeric eIF2 complex.

Two conserved strategies exist to control the activity of eIF4E. They involve either the binding of eIF4E to ligands that displace eIF4G, or the substitution of the active isoform of eIF4E with isoforms that have a defective eIF4G binding site. Both strategies have the same effect, namely that they prevent formation of the translation-promoting eIF4E:eIF4G complex. Interestingly, fungal cells appear to have often evolved one or the other solution but not both. Thus, baker's yeast contains two known ligands that in vitro can prevent binding of eIF4G (Altmann et al. 1997; Cosentino et al. 2000) and in vivo were shown to translationally control mRNAs bound by specific PUF proteins (Cridge et al. 2010). However, baker's yeast contains no inactive eIF4E isoform. In contrast, *Schizosaccharomyces pombe* contains an inactive eIF4E (Ptushkina et al. 2001), but does not contain known ligands able to displace eIF4G. As in mammals, TOR has been implicated in the regulation of eIF4E activity in fungi (Matsuo et al. 2005), although details of the regulatory mechanism are still virtually unexplored in the latter.

The second regulatory hub exerts control by regulating eIF2 activity, and is highly conserved among all eukaryotes. It impinges on a conserved phosphorylation site in the alpha subunit of this trimeric initiation factor. Phosphorylation at the conserved site turns this factor from a substrate of eIF2B into a tightly binding competitive inhibitor of this exchange factor (Hinnebusch 2005). Because eIF2B occurs in a substoichiometric ratio to eIF2 (von der Haar and McCarthy 2002),

phosphorylation of a small proportion of eIF2 is sufficient to disrupt the guanidine exchange cycle and to fully halt translation. At lower levels of phosphorylation, this mechanism can be used to fine-tune translational activity. In yeast, a single kinase, Gcn2, is known to phosphorylate eIF2 in response to amino acid starvation, and this represents one of the key steps of the general amino acid control pathway (Hinnebusch 2005), which is highly conserved among fungi.

Both eIF4E and eIF2 (via Gcn2) receive regulatory input from TOR, making this protein a master controller of translation. This master control extends to further levels, in particular to the control of ribosome biogenesis (Urban et al. 2007). Fungi in general, and in particular the unicellular fungi, are fast growing organisms, where a substantial proportion of protein synthesis is required to sustain growth (as opposed to replacing proteins lost due to protein turnover) (von der Haar 2008). The resulting high levels of activity can only be sustained if sufficient numbers of ribosomes are available, and these are the most expensive components of the translational machinery to produce. In consequence, at high growth rates, almost all ribosomes are active in translation. As the master controller of ribosome biogenesis, TOR therefore controls the overall capacity for translation as well.

mRNA Features that Determine the Efficiency of Protein Synthesis

Individual mRNAs compete for access to the total available translational capacity. How competitive they are is largely determined by their sequence, either because specific motifs alter interactions with translation factors, or because they recruit transacting proteins that then affect translational activity. The relative competitiveness of an mRNA is not static but shifts with changes in translation factor activities: typically, regulation of individual translation factors affects different messages in a nonlinear fashion, thus constantly changing the proportion of well-translated transcripts as translation factor activity is altered (Smirnova et al. 2005; Park et al. 2011b). This combinatorial relationship between translation factor activity and mRNA features is the basis for most of the translational control mechanisms that contribute to the adaptation of the proteome during stresses and development .

5′ UTR Length

On most mRNAs, initiation occurs by the canonical scanning mechanism as described above. Exceptions are mRNAs with very short and unstructured 5′-UTRs, which can initiate efficiently independently of the cap-binding complex. Thus, very short 5′-UTRs would render translation fully or partially independent of

regulatory input impinging on the cap-binding complex. Once a minimum length of the 5′-UTR has been surpassed, 5′-UTR length ceases to be a regulatory feature, since long 5′-UTRs in the absence of other inhibiting features like secondary structures or uORFs (see below) are not known to inhibit translation (Berthelot et al. 2004).

5′ UTR Structure

The requirement for helicase enzymes in translation initiation is brought about by the need to create an unstructured landing pad for the incoming 48S PIC, and to enable this complex to slide on the mRNA's 5′ leader during scanning. In consequence of this, the formation of secondary structures in the 5′ UTR almost invariably decreases translation initiation rates (an exception for this are so-called IRESs, specialised structures that have cap-independent ribosome recruitment activity, see below). Interestingly, secondary structures in the 5′ UTR are more inhibitory than similar structures in the ORF, since the scanning small subunit appears to have much poorer intrinsic helicase activity than the elongating, fully assembled ribosome.

The inhibitory effect of secondary structures is position dependent and increases with proximity to the cap, as well as increasing with thermodynamic stability (Koloteva et al. 1997). In mammalian cell free extracts or reconstituted in vitro systems, more strongly inhibitory structures confer a higher dependence of translation on eIF4A (Svitkin et al. 2001; Pestova and Kolupaeva 2002), and differential regulation of mRNAs with a higher secondary structure content may thus underlie the observed involvement of eIF4A in response to glucose depletion (Castelli et al. 2011) and adaptation to lithium stress (Montero-Lomeli et al. 2002) in yeast.

Start Codon Context

The ability of the scanning ribosome to recognise and accept start codons depends crucially on the sequence and context of a given start codon. Although AUG is by far the preferred start codon in all organisms, alternative codons like CUG are used as start sites of endogenous genes in baker's yeast (Chang and Wang 2004) and *Candida albicans* (Abramczyk et al. 2003), and are also active for the expression of reporter genes in *Neurospora* (Wei et al. 2013). Typically, in these near-cognate start codon cases, the efficiency of start codon recognition is orders of magnitude lower than in the case of AUG codons.

For both AUG codons and non-standard start sites, the efficiency of recognition is modulated by the nucleotides surrounding the start codon itself. In yeast, the best and worst contexts were experimentally determined as AAA and CGC immediately

upstream of the AUG, respectively (Chen et al. 2008), and correspondingly it was found that the three nucleotides upstream of start codons are under detectable selective pressure (Zur and Tuller 2013).

Upstream ORFs

A distinct regulatory feature of eukaryotic mRNAs are short (typically 3–10 codon long) open reading frames upstream of the main ORF. A typical regulatory role of these uORFs is to remove most of the ribosomes that initiate in a cap-dependent manner, thus leading to strongly repressed expression levels from uORF containing genes. If the uORF start codon is in a favourable context, almost all ribosomes will initiate translation on this start codon, and will then terminate when they encounter the stop codon. Following termination, the vast majority of ribosomes will be released from the mRNA. Thus, only a very small proportion of ribosomes that have either bypassed the uORF start codon, or that remain bound to the mRNA following termination and that restart scanning, will reach the start codon of the main ORF.

In rare cases, uORFs contain specific re-initiation-promoting sequences immediately upstream and/or downstream of their stop codon. In these cases, higher numbers of ribosomes (between 1 and 50 %) escape recycling and instead restart the scanning process. This process relies on contacts between the initiation factor eIF3, which remains bound to the elongating ribosome for at least the first five codons, and the specific re-initiation sequences (Szamecz et al. 2008; Munzarová et al. 2011).

One of the best studied cases of uORF-mediated regulation is that of the yeast transcriptional activator *GCN4* (Hinnebusch 2005). In this gene, a combination of multiple uORFs mediates intricate regulation that is inversely dependent on the availability of ternary complexes. The exact arrangement of uORFs varies between species, and ranges from two in *Neurospora* (Paluh et al. 1988) to four in *Saccharomyces cerevisiae* (Hinnebusch 2005), but GCN4 uORFs always render translation sensitive to Gcn2 kinase activity (see above). This kinase becomes activated, and indirectly inhibits eIF2 activity, in response to amino acid starvation as part of the fungal general amino acid control pathway.

On the GCN4 transcript, ribosomes initiate translation at the first, reinitiation-permissive uORF1, translate this and terminate translation. A high ($\sim$50 %) proportion of small subunits then resumes scanning, waiting to reacquire the eIF2:tRNA$_i^{Met}$:GTP ternary complex. If ternary complex availability is high (when Gcn2 is inactive), the reinitiating ribosomes reacquire this factor fast enough to become initiation competent before the encounter of subsequent uORFs, which are inhibitory for reinitiation. Following translation of these subsequent uORFs, ribosomes thus undergo the full round of recycling, are released from the mRNA, and never reach the main AUG. If availability of ternary complexes is low (when Gcn2 is active), many ribosomes scanning downstream of uORF1 only rebind the

ternary complex when they have moved past the inhibitory uORFs. In this case, they become competent for reinitiation at the *GCN4* AUG. In consequence, whereas translation of the majority of genes is reduced when Gcn2 becomes activated for example under starvation conditions, *GCN4* translation increases as the availability of ternary complexes decreases.

Although *GCN4* is the best studied example of fungal uORF-containing genes, it is far from being the only one. Sequence analyses revealed that at least 6 % of yeast genes express uORF containing mRNAs (Nagalakshmi et al. 2008), and uORF-dependent regulation has been implicated in a variety of biological pathways (Vilela and McCarthy 2003).

Codon Usage

As has likely become clear from the preceding paragraphs, control of translation is thought to reside mostly in the initiation phase. This is entirely expected from a highly processive mechanism where one initiation event on average produces one, or close to one, protein. Nevertheless, recent studies have shown that codon usage controls expression levels of some endogenous proteins in different fungi (Chan et al. 2012; Kemp et al. 2013; Zhou et al. 2013), and that this control of expression levels in turn controls various biological pathways from stress responses to circadian rhythms.

Experiments with codon usage variants of reporter genes in yeast have shown that the effect of codon usage on expression levels can be explained by the interference of slowly moving ribosomes with initiation of subsequent ribosomes (Chu et al. 2013). Thus, the combination of efficient initiation determinants with unfavourable codon usage can make translation-dependent on codon usage and ribosome speed, and can in theory make expression levels of such mRNAs responsive to elongation factor activity (Table 5.1).

IRESs

In some instances, translation initiation does not follow the canonical, cap-dependent and initiation factor-dependent pathways described above. Instead, ribosomes are recruited by sequence elements within the 5′-UTR, so-called internal ribosome entry sites or IRESs. Well-known examples of IRESs typically adopt complex secondary structures that directly interact with ribosomes and/or initiation factors (Thompson 2012), and such IRESs are widespread among viral mRNAs. IRESs typically allow initiation to proceed independently of some or all translation initiation factor activities, thus maintaining translational activity on selected, IRES-containing mRNAs under conditions where global translation is down-regulated.

The use of IRESs on cellular mRNAs is thought to be comparatively widespread in mammalian cells (Jackson 2013), although this question continues to be the cause

Table 5.1 Translation factors in *S. cerevisiae*

	Factor name	Obsolete names	Bacterial homolog	*S. cerevisiae* gene(s)	*S. cerevisiae* systematic	
Initiation	eIF1		IF3	*SUI1*	YNL244c	Promote 43S PIC assembly, scanning, and control start codon selection. eIF1A is also involved in the subunit joining process
	eIF1A	eIF-4C	IF1	*TIF11/FUN12*	YMR260c	
	eIF2α			*SUI2*	YJR007w	eIF2 binds GTP and the initiator tRNA and delivers it to the ribosomal A-site. Also involved in start codon selection
	eIF2*β*			*SUI3*	YPL237w	
	eIF2*γ*			*GCD11*	YER025w	
	eIF2Bα			*GCN3*	YKR026c	eIF2B is the GDP/GTP exchange factor (=guanidine exchange factor, GEF) for eIF2
	eIF2B*β*			*GCD7*	YLR291c	
	eIF2B*γ*			*GCD1*	YOR260w	
	eIF2B*δ*			*GCD2*	YGR083c	
	eIF2B*ε*			*GCD6*	YDR211w	
	eIF3a			*RPG1*	YBR079c	Promotes ternary complex and mRNA binding (43 and 48S PIC assembly), scanning and AUG recognition. Also controls translation termination and stop-codon read through
	eIF3b			*PRT1*	YOR361C	
	eIF3c			*NIP1*	YMR309c	
	eIF3 g			*TIF35*	YDR429c	
	eIF3i			*TIF34*	YMR146c	
	Hcr1 (eIF3j)			*HCR1*	YLR192c	Controls release factor activity during translation termination, as well as eIF3 binding and AUG recognition during initiation
	eIF4A			*TIF1, TIF2*	YKR059w, YJL138c	The eIF4A/eIF4B complex provides helicase activity to melt secondary structures in mRNAs
	eIF4B			*TIF3*	YPR163c	
	eIF4E			*CDC33*	YOL139c	The eIF4G:eIF4E complex (also referred to as eIF4F) binds to the mRNA 5′-cap and serves as a contact point for the initiating 40S ribosomal subunit
	eIF4G			*TIF4631, TIF4632*	YGR162w, YGL049c	
	eIF5			*TIF5*	YPR041w	Stimulates eIF2 GTPase activity during scanning and controls the GEF cycle
	eIF5B		IF2	*FUN12*	YAL035w	Catalyses subunit joining, together with eIF1A

(continued)

Table 5.1 (continued)

	Factor name	Obsolete names	Bacterial homolog	*S. cerevisiae* gene(s)	*S. cerevisiae* systematic	
Elongation	eEF1A	eEF1α	EF-Tu	*TEF1, TEF2*	YPR080w, YBR118w	Binds tRNAs and delivers them to the ribosomal A-site for the decoding process
	eEF1Bα	eEF1*β*	EF-Ts	*EFB1*	YAL003w	eIF1Bα/*γ* is the guanidine exchange factor for eIF1A
	eEF1B*γ*	eEF1*γ*		*TEF4*	YKL081w	
	eEF2		EF-G	*EFT1, EFT2*	YOR133w, YDR385w	Catalyses translocation from one codon to the next
	eEF3			*YEF3*	YLR249w	Catalyses release of the empty tRNA
	eIF5A	eIF-4D	EF-P	*ANB1, HYP2*	YJR047c, YEL034w	Specialised elongation factor required for efficient peptidyl transfer on multiple consecutive proline codons
Termination	eRF1		RF1, RF2	*SUP45*	YBR143c	Decodes the stop codon
	eRF3			*SUP35*	YDR172w	Binds GTP and aids eRF1 in stop codon decoding
	Rli1			*RLI1*	YDR091c	Controls eRF1-mediated peptidyl hydrolysis activity and catalyses subunit separation following termination

for debate (Kozak 2005). In fungi, only very few sequences have been suggested to function as IRESs. In baker's yeast, the *URE2* gene can initiate translation in a cap-independent manner (Komar et al. 2003), and genome-wide surveys identified multiple putative IRES elements in *Pichia pastoris* (Liang et al. 2012).

The Cell-Wide Network of Translation

One of the hallmarks of translation in vivo is that it relies on limiting cellular resources. Protein synthesis itself is an energy-hungry process that consumes at least five NTP in each elongation cycle (two for aminoacylation of the tRNA, and one each for codon decoding, ribosome translocation and empty tRNA release). Energy consumption in vivo is even higher due to the presence of unproductive processes, and has been estimated as 7 NTP per amino acid in animals (Aoyagi et al. 1988). Moreover, the biosynthesis of ribosomes requires additional energy investment for transcription of the ribosomal RNAs and the complex ribosomal biosynthesis process.

Under fast growth conditions, the majority of on-going protein synthesis is required to sustain cell division, rather than counteracting protein turnover (von der Haar 2008), so that the need for protein synthesis is directly linked to cellular growth rates. Several studies based on computational models of the cell-wide translational apparatus in baker's yeast have concluded that one of the major factors limiting growth is the availability of ribosomes (Chu and von der Haar 2012; Shah et al. 2013), and in consequence the translation systems of fast growing yeasts and fungi are likely to have evolved to allow production of the required proteome with a minimal number of ribosomes. Features of the transcriptome that allow such optimisation for ribosome usage include (i) selection for fast codons on highly expressed genes, which means that ribosomes are freed up faster for translation of the next message, and (ii) the avoidance of ribosomal traffic jams (which would unnecessarily block ribosomes) by expressing sufficient numbers of mRNAs, so that ribosome density on an individual message does not need to be too high.

One of the consequences of the limited availability of ribosomes, and the optimisation of ribosome usage, is that in fast growing fungi the majority of ribosomes is actively involved in translation. As we have discussed previously (Chu and von der Haar 2012), this means that it is difficult to regulate individual mRNAs in isolation. If an mRNA is down-regulated, the ribosomes engaged in its translation become available for translation of other mRNAs, and down-regulation of a large set of well-translated mRNAs would thus inevitably lead to the up-regulation of mRNAs which recruit ribosomes with lower than average efficiency. The need to prevent the uncontrolled propagation of regulatory events through the entire system via this or similar mechanisms may be one reason for the observed complexity of translational regulation, both in fungi and other organisms.

Acknowledgements This research was supported by the Czech Science Foundation Grant P305-11-0172 (to LV).

References

Abramczyk D, Tchórzewski M, Grankowski N (2003) Non-AUG translation initiation of mRNA encoding acidic ribosomal P2A protein in Candida albicans. Yeast 20:1045–1052. doi:10.1002/yea.1020

Acker MG, Shin B-S, Nanda JS et al (2009) Kinetic analysis of late steps of eukaryotic translation initiation. J Mol Biol 385:491–506. doi:10.1016/j.jmb.2008.10.029

Agris P, Vendeix F, Graham W (2007) tRNA's wobble decoding of the genome: 40 years of modification. J Mol Biol 366:1–13

Aitken CE, Lorsch JR (2012) A mechanistic overview of translation initiation in eukaryotes. Nat Struct Mol Biol 19:568–576. doi:10.1038/nsmb.2303

Altmann M, Schmitz N, Berset C, Trachsel H (1997) A novel inhibitor of cap-dependent translation initiation in yeast: p20 competes with eIF4G for binding to eIF4E. EMBO J 16:1114–1121. doi:10.1093/emboj/16.5.1114

Amrani N, Ghosh S, Mangus D a, Jacobson A (2008) Translation factors promote the formation of two states of the closed-loop mRNP. Nature 453:1276–1280. doi: 10.1038/nature06974

Andersen CBF, Becker T, Blau M et al (2006) Structure of eEF3 and the mechanism of transfer RNA release from the E-site. Nature 443:663–668. doi:10.1038/nature05126

Aoyagi Y, Tasaki I, Okumura J, Muramatsu T (1988) Energy cost of whole-body protein synthesis measured in vivo in chicks. Comp Biochem Physiol A: Comp Physiol 91:765–768

Becker T, Franckenberg S, Wickles S et al (2012) Structural basis of highly conserved ribosome recycling in eukaryotes and archaea. Nature 482:501–506. doi:10.1038/nature10829

Berthelot K, Muldoon M, Rajkowitsch L et al (2004) Dynamics and processivity of 40S ribosome scanning on mRNA in yeast. Mol Microbiol 51:987–1001. doi:10.1046/j.1365-2958.2003.03898.x

Beznoskova P, Cuchalová L, Wagner S et al (2013) Translation initiation factors eIF3 and HCR1 control translation termination and stop codon read-through in yeast cells. PLoS Genetics in press

Cannarrozzi G, Schraudolph NN, Faty M et al (2010) A role for codon order in translation dynamics. Cell 141:355–367. doi:10.1016/j.cell.2010.02.036

Castelli LM, Lui J, Campbell SG et al (2011) Glucose depletion inhibits translation initiation via eIF4A loss and subsequent 48S preinitiation complex accumulation, while the pentose phosphate pathway is coordinately up-regulated. Mol Biol Cell 22:3379–3393. doi:10.1091/mbc.E11-02-0153

Chan CTY, Pang YLJ, Deng W et al (2012) Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. Nat Commun 3:937. doi:10.1038/ncomms1938

Chang K-J, Wang C-C (2004) Translation initiation from a naturally occurring non-AUG codon in Saccharomyces cerevisiae. J Biol Chem 279:13778–13785. doi:10.1074/jbc.M311269200

Chen S-J, Lin G, Chang K-J et al (2008) Translational efficiency of a non-AUG initiation codon is significantly affected by its sequence context in yeast. J Biol Chem 283:3173–3180. doi:10.1074/jbc.M706968200

Christensen AK, Kahn LE, Bourne CM (1987) Circular polysomes predominate on the rough endoplasmic reticulum of somatotropes and mammotropes in the rat anterior pituitary. Am J Anat 178:1–10

Chu D, von der Haar T (2012) The architecture of eukaryotic translation. Nucleic Acids Res 40:10098–10106. doi:10.1093/nar/gks825

Chu D, Kazana E, Bellanger N et al (2013) Translation elongation can control translation initiation on eukaryotic mRNAs. EMBO Journal in press

Cosentino GP, Schmelzle T, Haghighat A et al (2000) Eap1p, a novel eukaryotic translation initiation factor 4E-associated protein in Saccharomyces cerevisiae. Mol Cell Biol 20:4604–4613

Cosson B, Couturier A, Chabelskaya S et al (2002) Poly (A)-binding protein acts in translation termination via eukaryotic release factor 3 interaction and does not influence [PSI(+)] propagation. Mol Cell Biol 22:3301–3315. doi:10.1128/MCB.22.10.3301

Cridge AG, Castelli LM, Smirnova JB et al (2010) Identifying eIF4E-binding protein translationally-controlled transcripts reveals links to mRNAs bound by specific PUF proteins. Nucleic Acids Res 38:8039–8050. doi:10.1093/nar/gkq686

Giegé R, Sissler M, Florentz C (1998) Universal rules and idiosyncratic features in tRNA identity. Nucleic Acids Res 26:5017–5035

Gutierrez E, Shin B-S, Woolstenhulme CJ et al (2013) eIF5A promotes translation of polyproline motifs. Mol Cell 51:35–45. doi:10.1016/j.molcel.2013.04.021

He H, von der Haar T, Singh CR et al (2003) The yeast eukaryotic initiation factor 4G (eIF4G) HEAT domain interacts with eIF1 and eIF5 and is involved in stringent AUG selection. Mol Cell Biol 23:5431–5445. doi:10.1128/MCB.23.15.5431

Hinnebusch AG (2005) Translational regulation of GCN4 and the general amino acid control of yeast. Annu Rev Microbiol 59:407–450

Hosoda N, Kobayashi T, Uchida N et al (2003) Translation termination factor eRF3 mediates mRNA decay through the regulation of deadenylation. J Biol Chem 278:38287–38291. doi:10.1074/jbc.C300300200

Jackson RJ (2013) The current status of vertebrate cellular mRNA IRESs. Cold Spring Harb Perspect Biol 5:a011569

Jackson RJ, Hellen CUT, Pestova T V (2012) Termination and post-termination events in eukaryotic translation. Advances in Protein Chemistry and Structural Biology, 1st ed. Elsevier Inc., pp 45–93

Jennings MD, Pavitt GD (2010) eIF5 has GDI activity necessary for translational control by eIF2 phosphorylation. Nature 465:378–381. doi:10.1038/nature09003.eIF5

Jivotovskaya AV, Valásek L, Hinnebusch AG, Nielsen KH (2006) Eukaryotic translation initiation factor 3 (eIF3) and eIF2 can promote mRNA binding to 40S subunits independently of eIF4G in yeast. Mol Cell Biol 26:1355–1372. doi:10.1128/MCB.26.4.1355-1372.2006

Kemp AJ, Betney R, Ciandrini L et al (2013) A yeast tRNA mutant that causes pseudohyphal growth exhibits reduced rates of CAG codon translation. Mol Microbiol 87:284–300. doi:10.1111/mmi.12096

Koloteva N, Müller PP, McCarthy JEG (1997) The position dependence of translational regulation via RNA-RNA and RNA-protein interactions in the 5′-untranslated region of eukaryotic mRNA is a function of the thermodynamic competence of 40 S ribosomes in translational initiation. J Biol Chem 272:16531–16539

Komar AA, Lesnik T, Cullin C et al (2003) Internal initiation drives the synthesis of Ure2 protein lacking the prion domain and affects [URE3] propagation in yeast cells. EMBO J 22:1199–1209. doi: 10.1093/emboj/cdg103

Kopeina GS, Afonina ZA, Gromova KV et al (2008) Step-wise formation of eukaryotic double-row polyribosomes and circular translation of polysomal mRNA. Nucleic Acids Research 36:2476–2488. doi: 10.1093/nar/gkm1177

Korneeva NL, Lamphear BJ, Hennigan FL et al (2001) Characterization of the two eIF4A-binding sites on human eIF4G-1. J Biol Chem 276:2872–2879. doi:10.1074/jbc.M006345200

Kovalchuke O, Ziehler J, Chakraburtty K (1995) Comparative analysis of ATPase of yeast elongation factor 3 and ATPase associated with Tetrahymena ribosomes. Biochimie 77:713–718

Kozak M (2005) A second look at cellular mRNA sequences said to function as internal ribosome entry sites. Nucleic Acids Res 33:6593–6602. doi:10.1093/nar/gki958

Le Sourd F, Boulben S, Le Bouffant R et al (2006) eEF1B: at the dawn of the 21st century. Biochim Biophys Acta 1759:13–31. doi:10.1016/j.bbaexp.2006.02.003

Liang S, Wang B, Pan L et al (2012) Comprehensive structural annotation of Pichia pastoris transcriptome and the response to various carbon sources using deep paired-end RNA sequencing. BMC Genom 13:738. doi:10.1186/1471-2164-13-738

Ling J, Reynolds NM, Ibba M (2009) Aminoacyl-tRNA synthesis and translational quality control. Annu Rev Microbiol 63:61–78

Matsuo R, Kubota H, Obata T et al (2005) The yeast eIF4E-associated protein Eap1p attenuates GCN4 translation upon TOR-inactivation. FEBS Lett 579:2433–2438. doi:10.1016/j.febslet.2005.03.043

Méthot N, Song MS, Sonenberg N (1996) A region rich in aspartic acid, arginine, tyrosine, and glycine (DRYG) mediates eukaryotic initiation factor 4B (eIF4B) self-association and interaction with eIF3. Mol Cell Biol 16:5328–5334

Miyazaki M, Kagiyama H (1990) Soluble factor requirements for the Tetrahymena peptide elongation system and the ribosomal ATPase as a counterpart of yeat elongation factor 3 (EF-3). J Biochem 108:1001–1008

Mohammed-Qureshi SS, Jennings MD, Pavitt GD (2008) Clues to the mechanism of action of eIF2B, the guanine-nucleotide-exchange factor for translation initiation. Biochem Soc Trans 36:658–664

Montero-Lomeli M, Morais BLB, Figueiredo DL et al (2002) The initiation factor eIF4A is involved in the response to lithium stress in Saccharomyces cerevisiae. J Biol Chem 277:21542–21548. doi:10.1074/jbc.M201977200

Munzarová V, Pánek J, Gunišová S et al (2011) Translation reinitiation relies on the interaction between eIF3a/TIF32 and progressively folded cis-acting mRNA elements preceding short uORFs. PLoS Genet 7:e1002137. doi:10.1371/journal.pgen.1002137

Nagalakshmi U, Wang Z, Waern K et al (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320:1344–1349. doi: 10.1126/science.1158441ET-2008/05/03

Paluh JL, Orbach MJ, Legerton TL, Yanofsky C (1988) The cross-pathway control gene of Neurospora crassa, cpc-1, encodes a protein similar to GCN4 of yeast and the DNA-binding domain of the oncogene v-jun-encoded protein. Proc Natl Acad Sci USA 85:3728–3732

Park E-H, Walker SE, Lee JM et al (2011a) Multiple elements in the eIF4G1 N-terminus promote assembly of eIF4G1•PABP mRNPs in vivo. EMBO J 30:302–316. doi:10.1038/emboj.2010.312

Park E-H, Zhang F, Warringer J et al (2011b) Depletion of eIF4G from yeast cells narrows the range of translational efficiencies genome-wide. BMC Genom 12:68. doi:10.1186/1471-2164-12-68

Pause a, Méthot N, Svitkin Y et al (1994) Dominant negative mutants of mammalian translation initiation factor eIF-4A define a critical role for eIF-4F in cap-dependent and cap-independent initiation of translation. EMBO J 13:1205–1215

Pestova TV, Kolupaeva VG (2002) The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. Genes Dev 16:2906–2922. doi:10.1101/gad.1020902

Pisareva VP, Pisarev AV, Komar AA et al (2008) Translation initiation on mammalian mRNAs with structured 5′UTRs requires DExH-box protein DHX29. Cell 135:1237–50. doi: 10.1016/j.cell.2008.10.037

Ptushkina M, Berthelot K, von der Haar T et al (2001) A second eIF4E protein in Schizosaccharomyces pombe has distinct eIF4G-binding properties. Nucleic Acids Res 29:4561–4569

Rodnina MV, Wintermeyer W (2009) Recent mechanistic insights into eukaryotic ribosomes. Curr Opin Cell Biol 21:435–443. doi:10.1016/j.ceb.2009.01.023

Saini AK, Nanda JS, Lorsch JR, Hinnebusch AG (2010) Regulatory elements in eIF1A control the fidelity of start codon selection by modulating tRNA(i)(Met) binding to the ribosome. Genes Dev 24:97–110. doi:10.1101/gad.1871910

Shah P, Ding Y, Niemczyk M et al (2013) Rate-limiting steps in yeast protein translation. Cell 153:1589–1601. doi:10.1016/j.cell.2013.05.049

Shoemaker CJ, Green R (2011) Kinetic analysis reveals the ordered coupling of translation termination and ribosome recycling in yeast. Proc Natl Acad Sci USA 108:E1392–E1398. doi:10.1073/pnas.1113956108

Smirnova JB, Selley JN, Sanchez-cabo F et al (2005) Global gene expression profiling reveals widespread yet distinctive translational responses to different eukaryotic translation initiation factor 2B-targeting stress pathways. Mol Cell Biol 25:9340–9349. doi:10.1128/MCB.25.21.9340

Spirin AS (2009) How does a scanning ribosomal particle move along the 5′-untranslated region of eukaryotic mRNA? Brownian Ratchet model. Biochemistry 48:10688–10692

Svitkin YV, Pause A, Haghighat A et al (2001) The requirement for eukaryotic initiation factor 4A (eIF4A) in translation is in direct proportion to the degree of mRNA 5′ secondary structure. RNA 7:382–394

Szamecz B, Rutkai E, Cuchalová L et al (2008) eIF3a cooperates with sequences 5′ of uORF1 to promote resumption of scanning by post-termination ribosomes for reinitiation on GCN4 mRNA. Genes Dev 22:2414–2425. doi:10.1101/gad.480508

Thompson SR (2012) Tricks an IRES uses to enslave ribosomes. Trends Microbiol 20:558–566. doi:10.1016/j.tim.2012.08.002

Topisirovic I, Svitkin YV, Sonenberg N, Shatkin AJ (2011) Cap and cap-binding proteins in the control of gene expression. Wires RNA 2:277–298

Urban J, Soulard A, Huber A et al (2007) Sch9 is a major target of TORC1 in Saccharomyces cerevisiae. Mol Cell 26:663–674. doi:10.1016/j.molcel.2007.04.020

Valášek LS (2012) "Ribozoomin"–translation initiation from the perspective of the ribosome-bound eukaryotic initiation factors (eIFs). Curr Protein Pept Sci 13:305–330

Vilela C, McCarthy JEG (2003) Regulation of fungal gene expression via short open reading frames in the mRNA 5′untranslated region. Mol Microbiol 49:859–867. doi:10.1046/j.1365-2958.2003.03622.x

von der Haar T, McCarthy JEG (2002) Intracellular translation initiation factor levels in Saccharomyces cerevisiae and their role in cap-complex function. Mol Microbiol 46:531–544

von der Haar T, Tuite MF (2007) Regulated translational bypass of stop codons in yeast. Trends Microbiol 15:78–86. doi:10.1016/j.tim.2006.12.002

von der Haar T (2008) A quantitative estimation of the global translational activity in logarithmically growing yeast cells. BMC Syst Biol 2:87. doi:10.1186/1752-0509-2-87

Vorhees RM, Ramakrishnan V (2013) Structural basis of the translational elongation cycle. Annu Rev Biochem 82:203–236

Walker SE, Zhou F, Mitchell SF et al (2013) Yeast eIF4B binds to the head of the 40S ribosomal subunit and promotes mRNA recruitment through its N-terminal and internal repeat domains. RNA 19:191–207. doi:10.1261/rna.035881.112

Wei J, Zhang Y, Ivanov IP, Sachs M (2013) The stringency of start codon selection in the filamentous fungus Neurospora crassa. J Biol Chem 288:9549–9562

Wells SE, Hillner PE, Vale RD, Sachs a B (1998) Circularization of mRNA by eukaryotic translation initiation factors. Mol Cell 2:135–140

Zhou M, Guo J, Cha J et al (2013) Non-optimal codon usage affects expression, structure and function of clock protein FRQ. Nature 495:111–115. doi:10.1038/nature11833

Zur H, Tuller T (2013) New universal rules of eukaryotic translation initiation fidelity. PLoS Comput Biol 9:e1003136. doi:10.1371/journal.pcbi.1003136

Chapter 6
mRNA Localization

Balaji T. Moorthy and Ralf-Peter Jansen

Abstract mRNA localization and localized translation is a common mechanism that contributes to cell polarity and cellular asymmetry. In metazoans, mRNA transport participates in embryonic axis determination and neuronal plasticity. Since the dissection of the mRNA localization process and its molecular machinery in higher eukaryotes has been found to be more arduous, unicellular eukaryotes like *Saccharomyces cerevisiae* or *Ustilago maydis* have become attractive models to study mRNA localization. Especially studies on cytoplasmic mRNA transport in *S. cerevisiae* have provided mechanistic insights as well as novel experimental approaches to mRNA localization, which have proven valuable for understanding similar processes in metazoans. In this review we will focus on mRNA localization of *ASH1* and other bud-localized mRNA in *S. cerevisiae* in order to provide a holistic view on mRNA localization. In addition, we will dwell on the mechanism and biological function of additional mRNA localization processes in budding yeast. Finally, we will compare mRNA transport in budding yeast with similar localization processes occurring in other fungi including the ascomycete *Candida albicans* and the basidiomycete *U. maydis*.

Contents

B. T. Moorthy · R.-P. Jansen (✉)
Interfaculty Institute of Biochemistry, Universität Tübingen, 72076 Tübingen, Germany
e-mail: ralf.jansen@uni-tuebingen.de

A. Sesma and T. von der Haar (eds.), *Fungal RNA Biology*,
DOI: 10.1007/978-3-319-05687-6_6,

Introduction to mRNA Localization

Localized mRNAs as source for targeted protein synthesis have been known for more than 20 years and were first described in oocytes of fruit flies, claw frogs, and marine invertebrates like tunicates where their localization and subsequent local translation are key to developmental processes such as tissue differentiation (for review see St Johnston 2005; Martin and Ephrussi 2009; St Johnston and Ahringer 2010). However, mRNA localization also occurs in several somatic cell types. In fibroblasts, targeting of *ß-actin* mRNA is essential for local actin polymerization and cell migration (Condeelis and Singer 2005). Neurons show an enrichment of more than 2,500 specific mRNAs in their axonal or dendritic processes (Cajigas et al. 2012) and local protein synthesis is required for axon guidance or synaptic plasticity (see Doyle and Kiebler 2011; Holt and Schuman 2013 for review).

The prevailing mechanism to achieve asymmetric mRNA distribution consists of active transport by motor proteins (Gagnon and Mowry 2011). Whereas cytoplasmic mRNA transport in oocytes and most somatic cells involves microtubule-dependent motor proteins such as kinesins or cytoplasmic dynein, myosin-driven transport along actin filaments has so far been reported for few cases only, including the transport of ß-actin mRNA in fibroblasts and of *ASH1* mRNA in budding yeast. Besides active transport, facilitated diffusion and anchoring can also result in local enrichment of specific mRNAs. The latter has been for example reported for *nanos* mRNA in *Drosophila* oocytes (Forrest and Gavis 2003).

Localized mRNAs are characterized by specific signals, so called localization elements (LEs) or zipcodes (Jambhekar and Derisi 2007). In most cases, LEs are located in the 3′ untranslated region (3′-UTR) of the mRNA but can also be found in the coding region (Jansen 2001; Jambhekar and Derisi 2007). LEs can dramatically vary in size, starting at 11 nucleotides and reaching several 100 nucleotides. This high variation already infers that there is little or no consensus sequence. Short LEs like the RNA trafficking signal in myelin basic protein mRNA (Munro et al. 1999) or a repetitive six nucleotide signal in *Xenopus laevis* Vg1 mRNA (Gautreau et al. 1997) might function on the nucleotide sequence level. However, most LEs function as structural elements (reviewed in Jambhekar and Derisi 2007). Although the structure of only few LEs, including the LE of Drosophila *K10* mRNA, have been determined to atomic detail (Bullock et al. 2010), secondary structure predictions or structural probing revealed that helical RNA structures are essential subelements of LEs.

The proteins recognizing these complex structural elements usually contain well known RNA-binding domains such as the RNA recognition motif (RRM; Maris et al. 2005) or the hnRNP K-homology (KH) domain (Valverde et al. 2008). As suggested by the complexity of most LEs, many of these zipcodes are not recognized by single RNA-binding proteins but can be simultaneously bound by several protein partners (Arn et al. 2003), generating large ribonucleoprotein (RNP) complexes that are transported to the target site.

mRNA Localization in *Saccharomyces cerevisiae*

In the fungal kingdom, the budding yeast *S. cerevisiae* has provided a pivotal model system to study mRNA localization. Its role as model for mRNA transport surpasses the fungal kingdom since many mechanistic insights into the molecular machinery of mRNA transport have initially been revealed in budding yeast (Bertrand et al. 1998; Böhl et al. 2000; Shen et al. 2010). Most mechanistic insights were derived from the analysis of a single mRNA, *ASH1* that is delivered to the daughter cell during mitosis. In addition, studies on mRNA localization to mitochondria or other organelles in yeast have initiated or boosted the search for similar phenomena in mammalian cells (Marc et al. 2002).

Bud-Specific Localization of ASH1 mRNA

Polarized transport of mRNAs from the mother cell to the bud has been observed for more than 30 different mRNAs (Long et al. 1997; Takizawa et al. 2000; Shepard et al. 2003; Aronov et al. 2007). However, in no other case do we understand the biological function and mechanism of site-specific mRNA delivery better than in the case localization of *ASH1* to the daughter cell during mitosis.

ASH1

Haploid cells of *S. cerevisiae* exist in two mating types, “a” or “α”. The mating type is, however, not stable and can change after each cell division. This process, called mating-type switching is a genomic recombination event at the *MAT* locus (Herskowitz 1989; Nasmyth 1993). An endonuclease encoded by the *HO* gene initiates mating-type switching by introducing a double strand break at the *MAT* locus. This is repaired using a silenced version of the *MAT* locus (*HMR* or *HML*) containing the genetic information of the opposite mating type (reviewed in Nasmyth 1993). Mating-type switching is initiated in G1 but only one cell of the previous mitotic division (the mother cell) switches its mating type whereas the daughter cell, which is derived from the bud cannot switch (Strathern and Herskowitz 1979).

Ash1p, a daughter-cell-specific transcriptional repressor of *HO* is key to the differential behavior of the two cells (Bobola et al. 1996; Sil and Herskowitz 1996). *ASH1* is activated during anaphase in mitosis and is expressed in both cells (Bobola et al. 1996; Sil and Herskowitz 1996). However, in late anaphase and prior to cell division the mRNA is transported to the distal tip of the daughter cell (Long et al. 1997). This ensures protein synthesis in the daughter cell only and thus daughter-cell-specific transcriptional inhibition of *HO*. Subsequently, rapid transcriptional shutdown and mRNA decay leads to quick removal of *ASH1* mRNA (Bobola et al. 1996) and phosphorylation of Ash1 protein by the Pho85p kinase initiates rapid degradation of the protein in daughter cells (McBride et al. 2001). Both mechanisms ensure that the protein is present in daughter cells only until the end of their first G1 cell cycle phase.

Localization of *ASH1* to the bud tip depends on localization elements within the mRNA. Four elements were identified in two parallel studies by the Singer and Nasmyth labs (Chartrand et al. 1999; Gonzalez et al. 1999). The size of these elements, termed E1, E2A, E2B, and E3 ranges from 115 nucleotides (E1) to 250 nucleotides (E2B). Three elements, E1, E2A, and E2B are positioned in the coding region, whereas the major part (100 nucleotides out of 118 in total) of the fourth element named E3 lies in the 3′UTR (Chartrand et al. 1999). The position of these elements is unusual since localization elements are generally found in the 3′-UTR (St Johnston 2005). Each *ASH1* localization element is capable of localizing the endogenous *ASH1* mRNA or a reporter mRNA to the bud, but tight association with the bud tip requires the presence of all LEs, indicating a cooperative function of the LEs (Chartrand et al. 1999; Gonzalez et al. 1999). In addition, the position of each individual element within the mRNAs is also important (Chartrand et al. 2002). When E1, E2A, and E2B elements within the coding region were destroyed by point mutations without disturbing the reading frame and functional copies of the same elements were placed in the 3′-UTR, this fully restored *ASH1* localization. However, the translational regulation of the mRNA (see "Khd1p and Puf6p: Translational control of ASH1 mRNA during transport") was disturbed, indicating that these structured RNA regions serve additional functions.

Secondary structure prediction of the localization elements and mutational analysis suggest that they share common features (Chartrand et al. 1999; Olivier et al. 2005). All elements are predicted to form bulged stem-loop structures and mutations disrupting the double strand that forms the more distal region of the stem severely impair their function. In addition, a CGA nucleotide triplet is conserved in all LEs. Localization elements similar to those in *ASH1* have also been identified in other bud-localized mRNAs (Olivier et al. 2005; Jambhekar et al. 2005).

The Repertoire of the Main Trans-Acting Factors for *ASH1* Localization

Localization of *ASH1* is an active process, first observed with the help of the MS2 RNA tagging system developed by Singer and colleagues (Bertrand et al. 1998).

This approach makes use of a specific interaction between the coat protein of the bacteriophage MS2 (MS2-CP) and RNA stem-loop structures (*MS2L*) in the viral RNA (Peabody and Lim 1996). If a fusion protein of MS2-CP and green fluorescent protein (GFP) is expressed in yeast, it will bind to coexpressed repetitive *MS2L* units introduced into the 3′-UTR of the RNA of choice (Urbinati and Long 2011). The resulting mRNPs can be visualized by fluorescence microscopy and analyzed regarding their directionality of transport and velocity of movement. Imaging of motile *ASH1-MS2L* particles provided the first direct evidence that mRNA localization of a specific mRNA relies on active transport (Bertrand et al. 1998). It requires a functional actin cytoskeleton (Long et al. 1997; Takizawa et al. 1997), a type V myosin motor protein (Jansen et al. 1996; Long et al. 1997; Takizawa et al. 1997), and five RNA-binding proteins (She2p, She3p, Loc1p, Khd1p, Puf6p) that fulfill different roles during *ASH1* localization, ranging from mRNP assembly via mRNA localization to translational inhibition (Böhl et al. 2000; Long et al. 2001; Irie et al. 2002; Gu et al. 2004).

She2p Among the five RNA-binding proteins, She2p has been studied in most detail. She2p binds to all four localization elements of *ASH1* mRNA (Böhl et al. 2000; Niessing et al. 2004). It also interacts with more than 50 additional mRNAs (Shepard et al. 2003; Oeffinger et al. 2007). The *SHE2* gene encodes a 28 kDa protein that can shuttle between cytoplasm and nucleus but is more abundant in the cytoplasm (Jansen et al. 1996; Kruse et al. 2002; Shen et al. 2009). Nuclear import depends on a noncanonical nuclear localization sequence at its carboxyterminus (Shen et al. 2009) whereas nuclear export occurs only after RNA-binding and uses the mRNA export pathway (Kruse et al. 2002). The crystal structure of She2p reveals a dimeric protein with α-helices as the prevalent structural elements of each subunit (Niessing et al. 2004). This dimer contains two RNA-binding domains of the unusual basic helical hairpin type. Two dimers assemble in solution in a head-to-head manner to form a tetrameric complex, thus generating four RNA-binding sites (Müller et al. 2009). The tetramer has an increased affinity to individual localization elements and a mutation in the dimer–dimer interface reduces the binding to individual localization elements up to 30fold (Müller et al. 2009). A characteristic structural feature of She2p is a short α-helix that protrudes at right angle from the remaining helices. This helix plays a specific role in selective binding to LEs and for interaction with She3p (Müller et al. 2011). A recent report suggests that cytoplasmic She2p can also bind to membranes (Genz et al. 2013). This might be relevant during cotransport of mRNAs encoding membrane proteins and endoplasmic reticulum (Schmid et al. 2006; Fundakowski et al. 2012). She2p preferentially interacts with highly curved membranes reminiscent of tubular ER structures but the mode of interaction is currently unknown (Genz et al. 2013).

She3p Like She2p, She3p was identified in a mutant screen for proteins affecting asymmetric *HO* expression and mutations in *SHE3* show a very strong defect in *ASH1* localization (Jansen et al. 1996; Long et al. 1997). Early studies suggested that this protein serves as linker between She2p and the myosin Myo4p since its amino terminal half (amino acids 1–234), which is predicted to form a

coiled-coil structure binds to the Myo4p and its carboxyterminus interacts with She2p (Böhl et al. 2000). However, later reports demonstrated that She3p can also directly bind *ASH1* LEs, forming a stable trimeric complex with She2p and individual localization elements in vitro (Müller et al. 2011). Binding to RNA occurs within a 20 amino acid region in the carboxyterminal domain (Müller et al. 2011) and mutations in this region affect *ASH1* localization (Landers et al. 2009).

Myo4p Myo4p is one of the two class V myosins present in *S. cerevisiae*. Class V myosins are usually dimeric actin-binding proteins that work as processive motors, i.e., they can move long distances along microfilaments without pausing or falling off (Sellers and Veigel 2006). However, both yeast myosins are nonprocessive in the absence of cargo (Reck-Peterson et al. 2001; Dunn et al. 2007). Although Myo4p shares several structural and functional features like a high duty cycle motor domain and a long lever arm with other class V members (Krementsova 2005), it is unique in other aspects. At its physiological concentration, unlike all other class V myosins studied so far, Myo4p is strictly monomeric (Heuck et al. 2007). This monomeric state is functionally related to the rod and globular tail domains of Myo4p since its motor domain, if fused to the rod and tail of mouse myosin V allows processive movement of the hybrid (Krementsova 2005). Myo4p stably associates with She3p via two distinct binding regions, the C-terminal tail and a coiled-coil domain containing fragment preceding the tail region (Heuck et al. 2007).

Long distance movement of Myo4p requires protein- or RNP complex-mediated oligomerization. In a minimal transport complex with an RNA carrying one LE or the She2p RNA-binding protein two Myo4p molecules are present (Chung and Takizawa 2010; Krementsova et al. 2011). A reconstituted complex of tetrameric She2p and two Myo4p-She3p heterodimers was shown to be sufficient for sustained hand-over-hand movement in single particle processivity assays in the absence of RNA cargo (Krementsova et al. 2011). This was surprising since previous reports had suggested a role of RNA in promoting complex assembly and transport (Chung and Takizawa 2010; Müller et al. 2011). However, a recent report using in vitro motility assays with reconstituted complexes supports the older findings (Sladewski et al. 2013). Here, it was demonstrated that the presence of an RNA with localization elements significantly increases the frequency of RNP runs along actin cables (Sladewski et al. 2013).

Loc1p Loc1p is an abundant nucleolar protein that is required for localization of *ASH1* (Long et al. 2001). In addition to mRNA localization, it participates in the biogenesis and nuclear export of 60S ribosomal subunits (Urbinati et al. 2006). Loc1p was identified in a three-hybrid interaction assay as a protein that binds to the E1 and E3 elements of *ASH1* but it also associates with other endogenous RNAs, preferentially to double-stranded RNA motifs (Long et al. 2001). Its specific recruitment to *ASH1* requires nuclear She2p (see "The current model for ASH1 mRNP assembly and localization"). Loc1p is required for proper localization (Long et al. 2001) and translational regulation of *ASH1* (Komili et al. 2007; Du et al. 2008) but its molecular function is not well understood. Three functions

have been suggested. Firstly, it might help to load the translational repressor Puf6p (see "Khd1p and Puf6p: Translational control of ASH1 mRNA during transport") onto *ASH1* mRNA (Shen et al. 2009), which could explain the translation repression defect of *loc1*Δ cells. Loc1p might also indirectly influence *ASH1* translation via assembly of specialized ribosomes that translate *ASH1* mRNA (Komili et al. 2007). Finally, it has been proposed to stabilize a nuclear complex of She2p and *ASH1* before export of the mRNP occurs (Niedner et al. 2013).

Khd1p and Puf6p: Translational Control of *ASH1* mRNA during Transport

In order to achieve local translation, the translation machinery needs to be blocked during cytoplasmic RNA transport (Besse and Ephrussi 2008). Two proteins, Khd1p and Puf6p directly participate in this regulation step in case of *ASH1*. Khd1p (also called Hek2p) belongs to the family of RNA-binding proteins containing heterologous nuclear ribonucleoprotein K (hnRNP K)-like homology (KH) domains (Irie et al. 2002; Valverde et al. 2008). It can directly bind to the E1 RNA localization element within *ASH1* and this binding leads to translational silencing due to interaction of Khd1p with eIF4G1 (Paquin et al. 2007), a major translation initiation factor (see Chap. 5). Binding of Khd1p to a carboxyterminal domain of eIF4G1 appears to counteract the otherwise positive effect of eIF4G1 on *ASH1* translation (Paquin et al. 2007). Translation inhibition is overridden by phosphorylation of Khd1p at several serine and threonine residues in the carboxyterminal 49 amino acids. This phosphorylation is mediated by the yeast casein kinase Yck1p and results in release of Khd1p from *ASH1* mRNA (Paquin et al. 2007). The cellular function of Khd1p by far exceeds translational control of *ASH1* mRNA since it has been reported to bind to more than 1,000 mRNAs, among them 12 other localized mRNAs (Hasegawa et al. 2008). Bioinformatics revealed a potential recognition motif, a repetitive CNN triplet (Hasegawa et al. 2008), which is unrelated to the CGA triplet that is part of the localization element (see "ASH1"). Although this motif is not present in *ASH1* E1, an array of multiple CNN motifs can be found 5′ of the E1 element. In contrast to its role in *ASH1* localization, the function of Khd1p is less clear for its other targets. *khd1*Δ cells show no translational misregulation of other localized mRNAs. Furthermore, Khd1p seems to be important for the stability of some of these mRNAs like *MTL1* without directly influencing their translation (Hasegawa et al. 2008).

The second translational regulator of *ASH1*, Puf6p was identified by copurification with She2p (Gu et al. 2004). Puf6p belongs to the family of pumilio-like RNA-binding proteins and is one of the six pumilio-like factors found in *S. cerevisiae* (Gerber et al. 2004; Gu et al. 2004). It is a shuttling protein like She2p but is highly abundant in the nucleolus where it participates in 60S ribosomal subunit biogenesis (Li et al. 2009). Within *ASH1*, it binds in the E3 element to a region containing a PUF consensus tetranucleotide UUGU (Gu et al. 2004). Binding of

Puf6p to *ASH1* blocks the assembly of 80S ribosomes at the start codon by inhibiting the recruitment of the 60S ribosomal subunit (Deng et al. 2008). Inhibition is caused by Puf6p interaction with the translation initiation factor eIF5B, which is essential for 80S assembly (Deng et al. 2008; see Chap. 5). Similar to the regulation of Khd1p, binding of Puf6p to *ASH1* is reduced by casein kinase mediated phosphorylation of the protein. In contrast to Yck1p that phosphorylates Khd1p, Puf6p phosphorylation is mediated by casein kinase 2 (Deng et al. 2008).

The Current Model for *ASH1* mRNP Assembly and Localization

The assembly of a localization-competent *ASH1* mRNP complex starts during transcription of *ASH1* by binding of nuclear She2p to the nascent *ASH1* transcript (Fig. 6.1). Nuclear import of She2p occurs in its monomeric form and requires binding to the yeast importin α-Srp1p (Shen et al. 2009). Since stable interaction with localization elements requires dimeric or tetrameric She2p (Müller et al. 2009), it must be assumed that the oligomeric form is generated after nuclear import or upon binding to RNA. The initial step in forming the *ASH1* mRNP is the association of She2p with elongating RNA polymerase II (Pol II), which requires the transcription elongation factors Spt4p and Spt5p (Shen et al. 2010). She2p can then be transferred to transcripts emerging from Pol II. Although RNA binding has been observed even to transcripts without localization elements, a more stable association is mediated by LEs in the emerging transcripts (Shen et al. 2010). Stabilization of the She2p-*ASH1*LE interaction is mediated by recruitment of Loc1p into the complex (Niedner et al. 2013). Since this protein resides in the nucleolus, a transfer of She2p-*ASH1* to the nucleolus has been proposed (Jellbauer and Jansen 2008). Supporting this hypothesis, nucleolar accumulation of She2p and *ASH1* mRNA has been observed when mRNA export from the nucleus is blocked (Du et al. 2008). Coimmunoprecipitation experiments suggest that nucleolar She2p might also recruit the translational repressor Puf6p into the complex (Shen et al. 2009). However, recent in vitro experiments on the assembly of She2p, Loc1p and Puf6p with an *ASH1* LE that includes the Puf6p binding site could not support these findings (Niedner et al. 2013). The premature mRNP formed in the nucleolus is subsequently exported into the cytoplasm via the Mex67p/TAP mRNA export pathway (Kruse et al. 2002; Du et al. 2008). It is not yet known whether the second translational inhibitor, Khd1p assembles into this complex before or after export to the cytoplasm. However, during or shortly after export a rearrangement within the mRNP occurs. Loc1p appears to leave the complex since it is not detectable in the cytoplasm. New data suggest that it is replaced by inclusion of She3p into the complex (Niedner et al. 2013). This rearrangement might also involve specific proteins of the nuclear pore. Deletion of the gene encoding the nonessential nucleoporin Nup60p has a specific effect on localized mRNAs compared to other transcripts (Powrie et al. 2010). Localized mRNAs including *ASH1* are inefficiently exported and their localization is impaired, possibly by an inefficient integration of She3p into the mRNP.

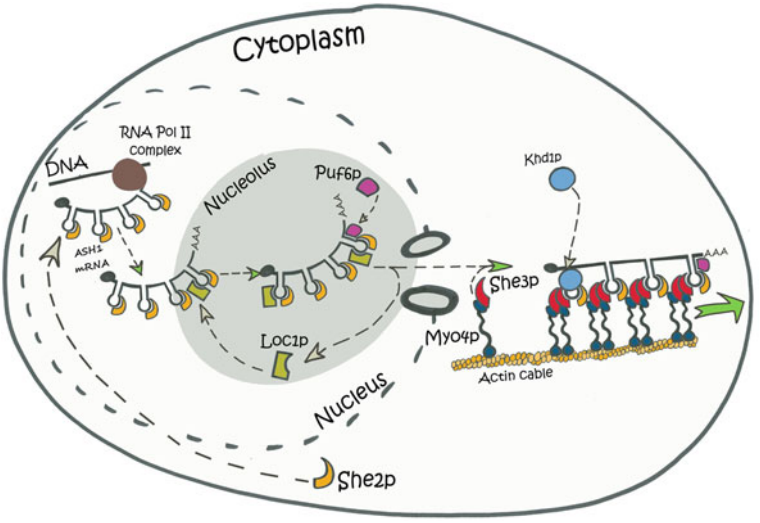


Fig. 6.1 The current model for ASH1 mRNP assembly and localization—assembly of the *ASH1* mRNP starts by nuclear import of She2p and its cotranscriptional recruitment to the four stem-loop localization elements of *ASH1*. This premature mRNP seems to be targeted by recruitment of Loc1p (which binds to the first and last localization element, E1 and E3) to the nucleolus where it recruits the translation repressor Puf6p to the E3 element. During—or directly after—export from the nucleus, Loc1p dissociates and is replaced by cytoplasmic She3p, which results in a stable She2p-She3p-*ASH1* complex. She3p is associated with monomeric Myo4p and binding of She3p ensures recruitment of multiple motors to *ASH1* mRNA. After binding of the second translational repressor Khd1p, the mature mRNP moves along actin cables to the tip of the daughter cell. For details see main text. Image by courtesy of Filipa Serra de Silva

Cotranscriptional recruitment and remodeling during RNP maturation is a common theme and has been observed for a number of localized mRNAs in other eukaryotes (Martin and Ephrussi 2009). For example, the ß-actin LE binding protein ZBP1 shuttles between cytoplasm and nucleus and can be detected at the site of ß-actin transcription (Oleynikov and Singer 2003). The exon junction complex, a multiprotein complex that is delivered to spliced mRNAs upstream of the exon–exon junction during (cotranscriptional) splicing is essential for Drosophila *oskar* mRNA localization (Hachet and Ephrussi 2001). Localization of Vg1 mRNA in *Xenopus* ooocytes involves several ordered remodeling steps that lead to a rearrangement of RNA-binding proteins within the mRNP (Kress et al. 2004). The mature complex finally recruits microtubule-dependent motor proteins.

In yeast, the recruitment of She3p into the premature mRNP complex does not only result in a stable trimeric complex with She2p and *ASH1* mRNA but also recruits the Myo4p myosin into the complex. Complex assembly can induce dimerization of the monomeric Myo4p at each LE, which is required for continuous transport of the mRNP (Chung and Takizawa 2010). *ASH1* mRNA with its four localization elements is able to recruit several transport complexes in vitro although not all elements are equally competent for recruitment (Sladewski et al. 2013).

Thus, the presence of multiple localization elements might be a specific feature of localized yeast mRNAs that are transported by a myosin that needs cargo-mediated oligomerization. The fully assembled, mature complex is transported along the actin cytoskeleton to the bud cortex. Single molecule assays with in vitro-assembled complexes revealed that these mRNP complexes are optimized for transport along actin bundles or cables (Sladewski et al. 2013) that stretch from the mother to the daughter cell (Moseley and Goode 2006). Once the complex is localized at the bud tip, its cargo is translationally activated by phosphorylation of Khd1p and Puf6p. Several reports also described that *ASH1* mRNA is anchored at the target site. This anchoring depends on translation of the mRNA (Gonzalez et al. 1999; Chartrand et al. 2002) and rearrangement of the mRNP (Landers et al. 2009). In addition, the exocyst protein complex located at the bud tip might be involved in anchoring since deletion of its nonessential component Sec3p results in mislocalization of *ASH1* mRNA (Aronov et al. 2007).

Additional Bud-Localized mRNAs Bound by She2p

Coimmunoprecipitation experiments demonstrated that She2p associates with more than 50 mRNAs (Shepard et al. 2003; Aronov et al. 2007; Oeffinger et al. 2007). The majority of them localize in a She-dependent manner to the bud (Shepard et al. 2003; Aronov et al. 2007). This She-dependent localization already suggested that these mRNAs share common features with *ASH1*. By analyzing the binding of fragments from these mRNAs to She2p and She3p, their LEs were identified (Jambhekar et al. 2005). Similar to *ASH1*, multiple LEs were found in the coding region of *WSC2*, *DNM1*, and *YLR434c* mRNAs, whereas only single LEs were observed in seven other localized mRNAs including *IST2* and *YMR171c/EAR1* (Olivier et al. 2005; Jambhekar et al. 2005). All LEs contain the conserved CGA triplet present in *ASH1* localization elements (see "ASH1").

Unlike the mitotically expressed *ASH1*, many of the other bud-localized mRNAs are either expressed throughout the cell cycle or between G1 and G2 phases (Shepard et al. 2003). This led to the question if the mRNAs localized by She2p are transported individually or together as composite mRNP granules containing different mRNAs. Therefore, an in vivo imaging approach was taken to follow differently tagged mRNAs (Lange et al. 2008). In addition to MS2 tagging (see "The repertoire of the main trans-acting factors for ASH1 localization"), a second RNA tagging system was implemented, based on the interaction of the boxB RNA stem-loop structures and an aminoterminal peptide from the antiterminator protein of bacteriophage lambda (Daigle and Ellenberg 2007; Lange et al. 2008). Lange et al. showed that boxB-tagged *WSC2* mRNA colocalizes with MS2L-tagged *IST2* during translocation into the bud whereas colocalization is not observed if *WSC2* carried a mutation in its LE. If tagged *ASH1* mRNA was coexpressed in the same cell cycle phase with any of the two mRNAs, it also colocalized with these during transport (Lange et al. 2008). These results imply that yeast localized mRNAs are

transported as large RNA-protein assemblies as it has been suggested for certain neuronal and *Drosophila* mRNAs (Kanai et al. 2004; Besse et al. 2009).

In case of the localized mRNAs that encode membrane or secreted proteins, a cotransport with tubular ER structures to the bud has been proposed (Schmid et al. 2006; Fundakowski et al. 2012). ER tubules that extend into the growing bud represent the structures from which cortical ER (also called plasma membrane-attached ER), a specialized ER at the plasma membrane derives (Du et al. 2004). The biogenesis of cortical ER depends on a large set of proteins including Myo4p and She3p (Estrada et al. 2003; Du et al. 2004). Many of the proteins involved in cortical ER biogenesis are also required for localization of polarized mRNAs that are transported early during bud development (Fundakowski et al. 2012). In contrast, *ASH1* localization, which occurs during mitosis, is independent of ER tubule movement. As shown for one of these localized mRNAs, *WSC2*, ER-mRNA cotransport is independent of its translation. Instead, the She2 protein mediates RNA association with ER (Aronov et al. 2007). Recent observations indicate that She2p behaves like a peripheral ER membrane protein (Schmid et al. 2006; Genz et al. 2013) and that a direct interaction of She2p with membrane lipids contributes to the association of mRNAs with ER (Genz et al. 2013).

Most proteins encoded by localized yeast mRNAs are asymmetrically distributed and enriched in the yeast bud (Shepard et al. 2003; Aronov et al. 2007). The proteins are functionally diverse and include membrane proteins, extracellular proteins, kinases, small GTPases, and several polarity determination and secretion factors (POLs). However, it is not known why these proteins are synthesized from localized transcripts and whether mRNA localization to the bud is needed for their cellular function as it has been demonstrated for *ASH1*. Surprisingly, several examples suggest that in most cases mRNA localization is not essential and only supports asymmetric protein distribution (Shepard et al. 2003; Franz et al. 2007; Aronov et al. 2007). This has been nicely demonstrated for Ist2p, a multimembrane-spanning protein of the cortical ER (Fischer et al. 2009). Although it was originally suggested that mRNA localization determines bud-specific cortical enrichment of the protein (Takizawa et al. 2000), later studies showed that this localization is mainly achieved via a peptide signal that is encoded within the localization element of *IST2* mRNA (Franz et al. 2007). However, mutations affecting the mRNA localization without disturbing the peptide signal as well as loss of She2p reduced the bud-specific signal of Ist2p. Similar observations were made for several POL mRNAs that encode membrane-associated proteins lacking membrane anchors, e.g., *SRO7* or *SEC3* (Aronov et al. 2007).

She-Independent mRNA Localization in S. cerevisiae

Although most of our understanding of how mRNAs are asymmetrically sorted to the bud stems from studying the She-dependent mRNA transport, several other mechanism have been described.

mRNA Localization during Shmooing

POL mRNAs (see above) are localized to the site of cell growth not only during budding but also during another mode of cell polarization called shmooing (Gelin-Licht et al. 2012). Haploid yeast treated with the opposite mating factor form polarized plasma membrane extensions in the direction of the pheromone gradient. These extensions (called shmoos) are larger and more elongated projections than buds and are analogous to membrane extensions (i.e., lamellipodia) seen in higher eukaryotes. Local translation of polarization factors like Sro7p at the shmoo tip aids in rapid response to external cues like mating factors and helps steering shmoo growth (Gelin-Licht et al. 2012). Surprisingly, in contrast to *SRO7* localization during budding, its targeting to the shmoo tip requires Myo4p but does not involve She2p or She3p. Instead, it depends on a large multi-KH domain protein, Scp160p (Wintersberger et al. 1995), which directly or indirectly binds to Myo4p (Gelin-Licht et al. 2012). Binding of POL mRNAs to Scp160p increases during pheromone response. Furthermore, deletion of *SCP160* reduces efficiency to properly respond to pheromone gradients, suggesting that POL mRNAs localization is important during mating response and that it is mediated by a different mechanism than during budding.

Daughter-Cell-Specific Enhancement of Translation by Kap104p

The nuclear import factor Kap104p belongs to the importin ß (or karyopherin ß) family of nuclear transport factors (Mosammaparast and Pemberton 2004). It is involved in the RNA release and reimport of two shuttling mRNA-binding proteins, Nab2p and Nab4p (Aitchison et al. 1996; van den Bogaart et al. 2009). Kap104p is a key factor in a novel mechanism of local translation regulation in the bud. A Kap104p-GFP fusion protein accumulates in areas of polarized growth like the bud tip in early mitosis (van den Bogaart et al. 2009). Using tetracystein-coupled reporter proteins that allow visualization of the translation site in situ (Gaietta et al. 2002), Poolman and colleagues could show that the site of Kap104p accumulation coincides with areas of enhanced translation (van den Bogaart et al. 2009). The suggested mechanism involves a more efficient release of Nap2p/Nap4p from associated mRNAs at sites of abundant Kap104p and a faster import of Nap2p and Nab4p into the nucleus of the presumptive daughter cell. Thus, the daughter cell's cytoplasm will be faster depleted of Nab2p and Nab4p, allowing increased translation of mRNAs in the future daughter cell.

mRNA Localization to the Yeast Mother Cell: ABP140

In contrast to all other mRNAs described above, *ABP140* mRNA is transported into the opposite direction, to the mother cell pole opposite of the bud (Kilchert and Spang 2011). Abp140p is an actin- and tRNA-binding protein with its

actin-binding domain located at the aminoterminal 17 amino acids (Riedl et al. 2008; D'Silva et al. 2011). It has been used as a marker for actin cables, which are bundles of actin filaments that cross the bud neck between mother cell and bud. Localization of *ABP140* mRNA to the mother cell requires translation of the aminoterminus with the actin-binding domain. It was suggested that this domain, upon translation and emergence from the ribosomal exit site, binds actin filaments and locks the nascent chain complex including the mRNA onto filamentous actin (Kilchert and Spang 2011). Since actin filaments polymerize in a formin-dependent manner at the bud tip, the cables extend from the bud into the mother cell (Yang and Pon 2002). This retrograde flow of actin filaments ensures trafficking of the translated *ABP140*-ribosome-nascent chain complex towards the distant mother cell pole. The cotranslational mRNA targeting mechanism observed for *ABP140* is not unique but adds to several known examples including mammalian Dia1 mRNA (Liao et al. 2011) or signal recognition particle-dependent nascent chain targeting to the ER (see "mRNA localization to the endoplasmic reticulum").

Localization of mRNAs to Organelles in S. cerevisiae

A large number of messenger RNAs encoding mitochondrial proteins or proteins translocated into the endoplasmic reticulum (ER) are enriched at the respective organelle membranes (Kraut-Cohen and Gerst 2010; Devaux et al. 2010; Hermesh and Jansen 2013). Most of these mRNAs are associated with ribosomes, indicating their translation. It therefore appears that such mRNAs are translated at or in close vicinity of the target organelle and that localized translation of mRNAs contributes to protein sorting (Weis et al. 2013).

mRNA Localization to the Endoplasmic Reticulum

Targeting of proteins to the endoplasmic reticulum in general involves the signal recognition particle (SRP) that pauses translation of proteins destined for the ER and delivers the nascent chain/ribosome/mRNA complex to the ER membrane (Blobel et al. 1979). Localization of mRNAs in these complexes thus occurs cotransitionally and depends on signal sequences in the translated protein but is independent of cis-acting signals within the RNA. However, recent studies of mRNAs associated with ER suggest that alternative pathways involving RNA-binding proteins at the ER membrane might also exist (Kraut-Cohen and Gerst 2010; Hermesh and Jansen 2013).

Diehn and colleagues used microarray-based identification of mRNAs to assess in a transcriptome-wide analysis the subset of mRNAs that are translated by polysomes bound to ER in contrast to cytoplasmic polysomes (Diehn et al. 2000, 2006). Besides mRNAs encoding secreted, membrane or ER proteins, they also

showed that mRNAs coding for proteins that lack signal sequences are enriched on ER-bound polysomes. Nicchitta and colleagues later extended these studies by carefully designed experiments that allowed them to separate ER-associated mRNAs from cytoplasmic mRNAs (Lerner et al. 2003). Their studies, performed on human cell lines revealed that ER-bound ribosomes also carry mRNAs coding for cytoplasmic proteins. In contrast to ribosomes translating ER-targeted proteins, these ribosomes detach after translation initiation from ER to complete translation in the cytoplasm (Reid and Nicchitta 2012). This suggests that translation at the cytoplasmic surface of the ER represents a default pathway followed by sorting of transcripts for further translation at the ER or in the cytosol.

However, these studies did not reveal the mechanism of mRNA targeting and the nature of the signal within the mRNAs. Studies on yeast mRNA targeting to the ER have recently shed more light on these issues. Gerst and colleagues used the MS2 tagging system to follow the distribution of several mRNA encoding membrane proteins (Kraut-Cohen et al. 2013). Nine out of 11 mRNAs analyzed colocalized to a high degree with ER. Surprisingly, as shown for two exemplary mRNAs, *USE1* and *SUC2*, targeting to ER is independent of their 3′-UTR and their translation (Kraut-Cohen et al. 2013). The degree of colocalization with ER coincides with their U-richness in the coding region, which is consistent with a previous observation that mRNAs coding for membrane proteins are rich in uracil (Prilusky and Bibi 2009). Thus, the high content of uracil might preselect by a yet unknown mechanism mRNAs coding for membrane proteins for ER association.

Alternative targeting mechanisms that function via the 3′-UTR have also been described. *PMP1* mRNA encodes the regulatory subunit of the plasma membrane H(+) ATPase and is a small protein of 40 amino acids. Due to its short peptide length it does not emerge enough from the ribosome exit tunnel to allow SRP binding and translation pausing. As shown by sedimentation analysis and fluorescence in situ hybridization (FISH), *PMP1* 3′-UTR is essential and sufficient for *PMP1* mRNA association with intracellular membranes (Loya et al. 2008). Interestingly, a repetitive UG-rich sequence within the 3′-UTR was found to contribute to membrane association, which suggests that in this case U-richness is of functional importance, too.

With the exception of She2p (see above), yeast RNA-binding proteins that recognize and target RNA to the ER have not yet been identified. Although specific proteins like Puf1p or Puf2p preferentially bind mRNAs encoding secreted or membrane proteins (Gerber et al. 2004; Hogan et al. 2008) and would therefore present excellent candidates for such factors, their deletion has no effect on the distribution of several mRNAs tested so far (Kraut-Cohen et al. 2013). However, the identification of membrane-spanning RNA-binding proteins in the ER of human cells has demonstrated the presence of such hypothetical factors. Here, the p180 protein, which contains a single transmembrane domain binds RNA with its large carboxyterminal lysine-rich domain and keeps mRNAs at the cytoplasmic face of the ER membrane (Cui et al. 2012). Although p180 is only found in

metazoans, its overexpression in yeast leads to an enhancement of ER-mRNA association (Becker et al. 1999), indicating that it can mediate RNA binding to ER in the absence of additional cell-specific factors.

mRNA Localization to Mitochondria

The vast majority of mitochondrial proteins is encoded by the nuclear genome and needs to be imported into mitochondria. Mitochondrial protein import generally requires an aminoterminal signal within the peptide sequence, which is recognized by the import machinery in the outer mitochondrial membrane (Neupert and Herrmann 2007). The common view that mitochondrial protein import occurs posttranslationally mainly stems from observations that in vitro synthesized mitochondrial precursor proteins can be imported into purified mitochondria (Neupert and Herrmann 2007). However, several studies also indicate the presence of an alternative, cotranslational pathway that requires localization of mRNAs encoding mitochondrial proteins to mitochondria. This pathway might be of special importance for proteins that upon full translation in the cytoplasm can aggregate and thus no longer be imported into mitochondria (Garcia et al. 2007). Since the experiments and conclusions leading to this alternative model of targeting have been the focus of recent reviews (Devaux et al. 2010; Weis et al. 2013), we will only briefly discuss this phenomenon.

A first global analysis of mRNAs associated with mitochondrial-bound versus cytoplasmic polysomes revealed that the mRNAs of more than 50 % of all mitochondrial proteins are overrepresented on mitochondrial-bound polysomes (Marc et al. 2002). A crucial question raised by these studies is if targeting of mRNAs coding for mitochondrial proteins (mMP) requires translation of the signal sequence within the encoded protein (cotranslational targeting) or if mMPs are targeted independent of translation via specific RNA-binding proteins. Studies on three mMPs, *ATM1*, *ATP2*, and *OXA1* suggested that the 3′-UTR is required for proper mitochondrial targeting or their association with mitochondria-bound polysomes (Corral-Debrinski et al. 2000; Gadir et al. 2011), suggesting that cis-acting signals are needed for their localization. Targeting of a large group of mMPs involves the RNA-binding protein Puf3p (Saint-Georges et al. 2008) and most of these mRNAs contain a UUGU Puf-binding consensus site. Since mutations in this site reduce colocalization of mMPs with mitochondria, at least a subfraction of mMPs might be targeted via their binding partner Puf3p (Saint-Georges et al. 2008).

In contrast to the above other studies suggested a need of translation for targeting of mMPs (Eliyahu et al. 2010; Garcia et al. 2010). Arava and colleagues reported that the mitochondrial import receptor Tom20p, which functions during the initial recognition of import signals in mitochondrial proteins (Neupert and Herrmann 2007) is required for proper association of mitochondria with several but not all mMP (Eliyahu et al. 2010), indicating that translation is vital for mMP

targeting. Interestingly, a further deletion of *PUF3* in *tom20Δ* cells exacerbated the defect. Thus, targeting via RNA-based signals and protein-based targeting might work in combination to achieve optimal protein delivery to mitochondria. Concomitantly, Gerst and colleagues found that colocalization of MS2L-tagged mMPs with mitochondria is strongly reduced if their 3′-UTR is deleted in combination with a mutation of the AUG codon (Gadir et al. 2011). In this case, colocalization is even further reduced if the mutant transcripts are expressed in *puf3*Δ cells. Thus, mRNA localization to yeast mitochondria seems to be achieved a combination of mRNA-dependent and cotranslational targeting.

mRNA Localization in Other Fungi

mRNA Localization in Candida albicans

In *S. cerevisiae*, *ASH1* does not only control HO expression but is also a transcriptional regulator of genes required for pseudohyphal growth (Chandarlapaty and Errede 1998). During pseudohyphal growth the cells become elongated, budding occurs synchronously in unipolar fashion, and the buds do not separate, producing a chain of cells which is called pseudohyphae (Gancedo 2001). Since pseudohyphal growth shares features of dimorphic differentiation processes like filamentation in other ascomycetes, e.g., *C. albicans,* this led to studies on *Candida* Ash1p function and *CaASH1* (*C. albicans ASH1*) mRNA localization. Mutations in *CaASH1* diminish filament formation and gene disruption results in reduced virulence of *C. albicans* in a mouse model (Inglis and Johnson 2002). Similar to Ash1p, CaAsh1p is preferentially localized to daughter cell nuclei in the budding-yeast form of *C. albicans* cells and to the hyphal tip cells in growing filaments. Thus, Ash1p "marks" newly formed cells and presumably directs a specialized transcriptional program in these cells. Asymmetric protein distribution is achieved by *CaASH1* mRNA localization to the daughter cell in the budding-yeast stage and to the hyphal tip during hyphal growth (Elson et al. 2009). Genomic screening for She proteins in *C. albicans* revealed a homolog of budding yeast She3p and a single class V myosin. Deletion of *CaSHE3* abolished *CaASH1* mRNA localization, indicating that a She-like transport mechanism is present in *C. albicans*. This was corroborated by findings that *CaASH1*, when expressed in *S. cerevisiae* is localized to the daughter cell (Münchow et al. 2002). Similar to She3p in *S. cerevisiae*, the Candida homolog can bind to *CaASH1*. Additional 40 mRNAs can be coimmunoprecipitated with CaShe3p (Elson et al. 2009). These encode cell wall or membrane proteins as well as transcription factors and thus represent a similar spectrum of mRNAs as in budding yeast. Most surprisingly, no *SHE2*-like gene has been identified in *C. albicans* (Elson et al. 2009). This is consistent with an analysis of several fungal genomes (Güldener et al. 2004) revealing that *SHE2*

is only present in the *Saccharomyces* clade of ascomycetes. It is possible that a yet unknown RNA-binding protein replaces She2p in *C. albicans* mRNA localization. Alternatively, CaShe3p (in contrast to its *S. cerevisae* counterpart) could provide sufficient specificity for long distance mRNAs transport in filaments.

mRNA Localization in Ustilago maydis

U. maydis is a basidiomycete that infects corn and causes smut disease (Doehlemann and Mendoza-Mendoza 2009). It can switch from yeast to filamentous forms after mating of two haploid cells. Filaments grow in a unipolar fashion, which depends on microtubule function. Mutants defective in the microtubule cytoskeleton or kinesin-1 show an aberrant bipolar filamentous growth pattern. Similar defects are seen in cells lacking the RNA-binding protein Rrm4 (Becht et al. 2006). Using a UV crosslinking and immunoprecipiation (CLIP) approach, Feldbrügge and colleagues identified a distinct set of Rrm4 mRNA targets sharing a CA-rich sequence motif (König et al. 2009). These mRNAs encode cytotopically related proteins such as translation or polarity factors. Rrm4 and the two mRNAs that were studied in more detail (*RHO3* and *UBI1*) show bidirectional movement along microtubules in filaments and colocalization of Rrm4 with target mRNAs during trafficking indicates the formation of mRNPs containing Rrm4 (König et al. 2009). The movement is similar to the one observed for dendritically targeted mRNAs in neurons and was nicknamed "sushi-belt" transport model (Doyle and Kiebler 2011). This model proposes that mRNAs in transported RNP particles can be discharged at several target sites along the track of movement and thus provide templates for localized translation at sites of demand. Several microtubule-dependent motor proteins participate in bidirectional transport in *U. maydis*. Among them are the kinesin Kin1 and the split dynein Dyn1/2, which appear to directly transport mRNPs and Kin3, which is required for dynein recycling (Baumann et al. 2012). Interestingly, the same set of motors also transports endosomes along the filament and Rrm4-containing mRNPs were shown to colocalize with shuttling endosomes (Baumann et al. 2012). The association of Rrm4 with endosomes is maintained in kinesin mutants, which suggests a piggyback type of RNA transport on endosomes in *U. maydis* (Baumann et al. 2012) and is reminiscent of another type of membrane-RNA cotransport, the ER-dependent localization of mRNAs in *S. cerevisiae* (Fundakowski et al. 2012).

Outlook

Although many of the mRNAs and RNA transport factors identified in fungi are not conserved in higher eukaryotes, fungal models might nevertheless allow important conclusions on the mechanisms of mRNA localization in more complex

organisms. Especially the understanding of the molecular details might reveal general principles and mechanisms of regulation, which could provide testable hypotheses for other model systems.

References

Aitchison JD, Blobel G, Rout MP (1996) Kap104p: a karyopherin involved in the nuclear transport of messenger RNA binding proteins. Science 274:624–627

Arn EA, Cha BJ, Theurkauf WE, Macdonald PM (2003) Recognition of a bicoid mRNA localization signal by a protein complex containing swallow, Nod, and RNA binding proteins. Dev Cell 4:41–51

Aronov S, Gelin-Licht R, Zipor G et al (2007) mRNAs encoding polarity and exocytosis factors are cotransported with the cortical endoplasmic reticulum to the incipient bud in *Saccharomyces cerevisiae*. Mol Cell Biol 27:3441–3455

Baumann S, Pohlmann T, Jungbluth M et al (2012) Kinesin-3 and dynein mediate microtubule-dependent co-transport of mRNPs and endosomes. J Cell Sci

Becht P, König J, Feldbrügge M (2006) The RNA-binding protein Rrm4 is essential for polarity in Ustilago maydis and shuttles along microtubules. J Cell Sci 119:4964–4973

Becker F, Block-Alper L, Nakamura G et al (1999) Expression of the 180-kD ribosome receptor induces membrane proliferation and increased secretory activity in yeast. J Cell Biol 146:273–284

Bertrand E, Chartrand P, Schaefer M et al (1998) Localization of ASH1 mRNA particles in living yeast. Mol Cell 2:437–445

Besse F, Ephrussi A (2008) Translational control of localized mRNAs: restricting protein synthesis in space and time. Nat Rev Mol Cell Biol 9:971–980

Besse F, López de Quinto S, Marchand V et al (2009) Drosophila PTB promotes formation of high-order RNP particles and represses oskar translation. Genes Dev 23:195–207

Blobel G, Walter P, Chang CN et al (1979) Translocation of proteins across membranes: the signal hypothesis and beyond. Symp Soc Exp Biol 33:9–36

Bobola N, Jansen RP, Shin TH, Nasmyth K (1996) Asymmetric accumulation of Ash1p in postanaphase nuclei depends on a myosin and restricts yeast mating-type switching to mother cells. Cell 84:699–709

Böhl F, Kruse C, Frank A et al (2000) She2p, a novel RNA-binding protein tethers ASH1 mRNA to the Myo4p myosin motor via She3p. EMBO J 19:5514–5524

Bullock SL, Ringel I, Ish-Horowicz D, Lukavsky PJ (2010) A'-form RNA helices are required for cytoplasmic mRNA transport in *Drosophila*. Nat Struct Mol Biol 17:703–709

Cajigas IJ, Tushev G, Will TJ et al (2012) The local transcriptome in the synaptic neuropil revealed by deep sequencing and high-resolution imaging. Neuron 74:453–466

Chandarlapaty S, Errede B (1998) Ash1, a daughter cell-specific protein, is required for pseudohyphal growth of *Saccharomyces cerevisiae*. Mol Cell Biol 18:2884–2891

Chartrand P, Meng XH, Huttelmaier S et al (2002) Asymmetric sorting of ash1p in yeast results from inhibition of translation by localization elements in the mRNA. Mol Cell 10:1319–1330

Chartrand P, Meng XH, Singer RH, Long RM (1999) Structural elements required for the localization of ASH1 mRNA and of a green fluorescent protein reporter particle in vivo. Curr Biol 9:333–338

Chung S, Takizawa PA (2010) Multiple Myo4 motors enhance ASH1 mRNA transport in *Saccharomyces cerevisiae*. J Cell Biol 189:755–767

Condeelis J, Singer RH (2005) How and why does beta-actin mRNA target? Biol Cell 97:97–110

Corral-Debrinski M, Blugeon C, Jacq C (2000) In yeast, the 3' untranslated region or the presequence of ATM1 is required for the exclusive localization of its mRNA to the vicinity of mitochondria. Mol Cell Biol 20:7881–7892

Cui XA, Zhang H, Palazzo AF (2012) p180 promotes the ribosome-independent localization of a subset of mRNA to the endoplasmic reticulum. PLoS Biol 10:e1001336
D'Silva S, Haider SJ, Phizicky EM (2011) A domain of the actin binding protein Abp140 is the yeast methyltransferase responsible for 3-methylcytidine modification in the tRNA anti-codon loop. RNA 17:1100–1110
Daigle N, Ellenberg J (2007) LambdaN-GFP: an RNA reporter system for live-cell imaging. Nat Meth 4:633–636
Deng Y, Singer RH, Gu W (2008) Translation of ASH1 mRNA is repressed by Puf6p-Fun12p/eIF5B interaction and released by CK2 phosphorylation. Genes Dev 22:1037–1050
Devaux FDR, Lelandais G, Garcia M et al (2010) Posttranscriptional control of mitochondrial biogenesis: spatio-temporal regulation of the protein import process. FEBS Lett 584:4273–4279
Diehn M, Bhattacharya R, Botstein D, Brown PO (2006) Genome-scale identification of membrane-associated human mRNAs. PLoS Genet 2:e11
Diehn M, Eisen MB, Botstein D, Brown PO (2000) Large-scale identification of secreted and membrane-associated gene products using DNA microarrays. Nat Genet 25:58–62
Doehlemann G, Mendoza-Mendoza A (2009) Ustilago maydis as a pathogen. Ann Rev Phytopath 47:423–445
Doyle M, Kiebler MA (2011) Mechanisms of dendritic mRNA transport and its role in synaptic tagging. EMBO J 30:3540–3552
Du T-G, Jellbauer S, Müller M et al (2008) Nuclear transit of the RNA-binding protein She2 is required for translational control of localized ASH1 mRNA. EMBO Rep 9:781–787
Du Y, Ferro-Novick S, Novick P (2004) Dynamics and inheritance of the endoplasmic reticulum. J Cell Sci 117:2871–2878
Dunn BD, Sakamoto T, Hong M-SS et al (2007) Myo4p is a monomeric myosin with motility uniquely adapted to transport mRNA. J Cell Biol 178:1193–1206
Eliyahu E, Pnueli L, Melamed D et al (2010) Tom20 mediates localization of mRNAs to mitochondria in a translation-dependent manner. Mol Cell Biol 30:284–294
Elson SL, Noble SM, Solis NV et al (2009) An RNA transport system in *Candida albicans* regulates hyphal morphology and invasive growth. PLoS Genet 5:e1000664
Estrada P, Kim J, Coleman J et al (2003) Myo4p and She3p are required for cortical ER inheritance in *Saccharomyces cerevisiae*. J Cell Biol 163:1255–1266
Fischer MA, Temmerman K, Ercan E et al (2009) Binding of plasma membrane lipids recruits the yeast integral membrane protein Ist2 to the cortical ER. Traffic 10:1084–1097
Forrest KM, Gavis ER (2003) Live imaging of endogenous RNA reveals a diffusion and entrapment mechanism for nanos mRNA localization in *Drosophila*. Curr Biol 13:1159–1168
Franz A, Maass K, Seedorf M (2007) A complex peptide-sorting signal, but no mRNA signal, is required for the Sec-independent transport of Ist2 from the yeast ER to the plasma membrane. FEBS Lett 581:401–405
Fundakowski J, Hermesh O, Jansen R-P (2012) Localization of a subset of yeast mRNAs depends on inheritance of endoplasmic reticulum. Traffic 13:1642–1652
Gadir N, Haim-Vilmovsky L, Kraut-Cohen J, Gerst JE (2011) Localization of mRNAs coding for mitochondrial proteins in the yeast *Saccharomyces cerevisiae*. RNA 17:1551–1565
Gagnon JA, Mowry KL (2011) Molecular motors: directing traffic during RNA localization. Crit Rev Biochem Mol Biol 46:229–239
Gaietta G, Deerinck TJ, Adams SR et al (2002) Multicolor and electron microscopic imaging of connexin trafficking. Science 296:503–507
Gancedo JM (2001) Control of pseudohyphae formation in *Saccharomyces cerevisiae*. FEMS Microbiol Rev 25:107–123
Garcia M, Darzacq X, Delaveau T et al (2007) Mitochondria-associated yeast mRNAs and the biogenesis of molecular complexes. Mol Biol Cell 18:362–368
Garcia M, Delaveau T, Goussard S, Jacq C (2010) Mitochondrial presequence and open reading frame mediate asymmetric localization of messenger RNA. EMBO Rep 11:285–291

Gautreau D, Cote CA, Mowry KL (1997) Two copies of a subelement from the Vg1 RNA localization sequence are sufficient to direct vegetal localization in Xenopus oocytes. Development 124:5013–5020

Gelin-Licht R, Paliwal S, Conlon P et al (2012) Scp160-dependent mRNA trafficking mediates pheromone gradient sensing and chemotropism in yeast. Cell Reports 1:483–494

Genz C, Fundakowski J, Hermesh O et al (2013) Association of the yeast RNA-binding protein She2p with the tubular endoplasmic reticulum depends on membrane curvature. J Biol Chem 288:32384–32393

Gerber AP, Herschlag D, Brown PO (2004) Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. PLoS Biol 2:E79

Gonzalez I, Buonomo SB, Nasmyth K, von Ahsen U (1999) ASH1 mRNA localization in yeast involves multiple secondary structural elements and Ash1 protein translation. Curr Biol 9:337–340

Gu W, Deng Y, Zenklusen D, Singer RH (2004) A new yeast PUF family protein, Puf6p, represses ASH1 mRNA translation and is required for its localization. Genes Dev 18:1452–1465

Güldener U, Münsterkötter M, Kastenmüller G et al (2004) CYGD: the comprehensive yeast genome database. Nucleic Acids Res 33:D364–D368

Hachet O, Ephrussi A (2001) Drosophila Y14 shuttles to the posterior of the oocyte and is required for oskar mRNA transport. Curr Biol 11:1666–1674

Hasegawa Y, Irie K, Gerber AP (2008) Distinct roles for Khd1p in the localization and expression of bud-localized mRNAs in yeast. RNA 14:2333–2347

Hermesh O, Jansen R-P (2013) Take the (RN)A-train: localization of mRNA to the endoplasmic reticulum. Biochim Biophys Acta 1833:2519–2525

Herskowitz I (1989) A regulatory hierarchy for cell specialization in yeast. Nature 342:749–757

Heuck A, Du T-G, Jellbauer S et al (2007) Monomeric myosin V uses two binding regions for the assembly of stable translocation complexes. Proc Natl Acad Sci USA 104:19778–19783

Hogan DJ, Riordan DP, Gerber AP et al (2008) Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. PLoS Biol 6:e255

Holt CE, Schuman EM (2013) The central dogma decentralized: new perspectives on RNA function and local translation in neurons. Neuron 80:648–657

Inglis DO, Johnson AD (2002) Ash1 protein, an asymmetrically localized transcriptional regulator, controls filamentous growth and virulence of *Candida albicans*. Mol Cell Biol 22:8669–8680

Irie K, Tadauchi T, Takizawa PA et al (2002) The Khd1 protein, which has three KH RNA-binding motifs, is required for proper localization of ASH1 mRNA in yeast. EMBO J 21:1158–1167

Jambhekar A, Derisi JL (2007) Cis-acting determinants of asymmetric, cytoplasmic RNA transport. RNA 13:625–642

Jambhekar A, McDermott K, Sorber K et al (2005) Unbiased selection of localization elements reveals cis-acting determinants of mRNA bud localization in *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA 102:18005–18010

Jansen RP (2001) mRNA localization: message on the move. Nat Rev Mol Cell Biol 2:247–256

Jansen RP, Dowzer C, Michaelis C et al (1996) Mother cell-specific HO expression in budding yeast depends on the unconventional myosin myo4p and other cytoplasmic proteins. Cell 84:687–697

Jellbauer S, Jansen R-P (2008) A putative function of the nucleolus in the assembly or maturation of specialized messenger ribonucleoprotein complexes. RNA Biol 5:225–229

Kanai Y, Dohmae N, Hirokawa N (2004) Kinesin transports RNA: isolation and characterization of an RNA-transporting granule. Neuron 43:513–525

Kilchert C, Spang A (2011) Cotranslational transport of ABP140 mRNA to the distal pole of *S. cerevisiae*. EMBO J 30:3567–3580

Komili S, Farny NG, Roth FP, Silver PA (2007) Functional specificity among ribosomal proteins regulates gene expression. Cell 131:557–571
König J, Baumann S, Koepke J et al (2009) The fungal RNA-binding protein Rrm4 mediates long-distance transport of ubi1 and rho3 mRNAs. EMBO J 28:1855–1866
Kraut-Cohen J, Afanasieva E, Haim-Vilmovsky L et al (2013) Translation- and SRP-independent mRNA targeting to the endoplasmic reticulum in the yeast *Saccharomyces cerevisiae*. Mol Biol Cell 24:3069–3084
Kraut-Cohen J, Gerst JE (2010) Addressing mRNAs to the ER: cis sequences act up! Trends Biochem Sci 35:459–469
Krementsova EB, Hodges AR, Lu H, Trybus KM (2005) Processivity of chimeric class V myosins. J Biol Chem 281:6079–6086
Krementsova EB, Hodges AR, Bookwalter CS et al (2011) Two single-headed myosin V motors bound to a tetrameric adapter protein form a processive complex. J Cell Biol 195:631–641
Kress TL, Yoon YJ, Mowry KL (2004) Nuclear RNP complex assembly initiates cytoplasmic RNA localization. J Cell Biol 165:203–211
Kruse C, Jaedicke A, Beaudouin J et al (2002) Ribonucleoprotein-dependent localization of the yeast class V myosin Myo4p. J Cell Biol 159:971–982
Landers SM, Gallas MR, Little J, Long RM (2009) She3p possesses a novel activity required for ASH1 mRNA localization in *Saccharomyces cerevisiae*. Eukaryot Cell 8:1072–1083
Lange S, Katayama Y, Schmid M et al (2008) Simultaneous transport of different localized mRNA species revealed by live-cell imaging. Traffic 9:1256–1267
Lerner RS, Seiser RM, Zheng T et al (2003) Partitioning and translation of mRNAs encoding soluble proteins on membrane-bound ribosomes. RNA 9:1123–1137
Li Z, Lee I, Moradi E et al (2009) Rational extension of the ribosome biogenesis pathway using network-guided genetics. PLoS Biol 7:e1000213–e1000213
Liao G, Ma X, Liu G (2011) An RNA-zipcode-independent mechanism that localizes dia1 mRNA to the perinuclear ER through interactions between dia1 nascent peptide and Rho-GTP. J Cell Sci 124:589–599
Long RM, Gu W, Meng X et al (2001) An exclusively nuclear RNA-binding protein affects asymmetric localization of ASH1 mRNA and Ash1p in yeast. J Cell Biol 153:307–318
Long RM, Singer RH, Meng X et al (1997) Mating type switching in yeast controlled by asymmetric localization of ASH1 mRNA. Science 277:383–387
Loya A, Pnueli L, Yosefzon Y et al (2008) The 3'-UTR mediates the cellular localization of an mRNA encoding a short plasma membrane protein. RNA 14:1352–1365
Marc P, Margeot A, Devaux F et al (2002) Genome-wide analysis of mRNAs targeted to yeast mitochondria. EMBO Rep 3:159–164
Maris C, Dominguez C, Allain FH-T (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. FEBS J 272:2118–2131
Martin KC, Ephrussi A (2009) mRNA localization: gene expression in the spatial dimension. Cell 136:719–730
McBride HJ, Sil A, Measday V et al (2001) The protein kinase Pho85 is required for asymmetric accumulation of the Ash1 protein in *Saccharomyces cerevisiae*. Mol Microbiol 42:345–353
Mosammaparast N, Pemberton LF (2004) Karyopherins: from nuclear-transport mediators to nuclear-function regulators. Trends Cell Biol 14:547–556
Moseley JB, Goode BL (2006) The yeast actin cytoskeleton: from cellular function to biochemical mechanism. Microbiol Mol Biol Rev 70:605–645
Munro TP, Magee RJ, Kidd GJ et al (1999) Mutational analysis of a heterogeneous nuclear ribonucleoprotein A2 response element for RNA trafficking. J Biol Chem 274:34389–34395
Müller M, Heym RG, Mayer A et al (2011) A cytoplasmic complex mediates specific mRNA recognition and localization in yeast. PLoS Biol 9:e1000611
Müller M, Richter K, Heuck A et al (2009) Formation of She2p tetramers is required for mRNA binding, mRNP assembly, and localization. RNA 15:2002–2012
Münchow S, Ferring D, Kahlina K, Jansen R-P (2002) Characterization of *Candida albicans* ASH1 in *Saccharomyces cerevisiae*. Curr Genet 41:73–81

Nasmyth K (1993) Regulating the HO endonuclease in yeast. Curr Opin Genet Dev 3:286–294
Neupert W, Herrmann JM (2007) Translocation of proteins into mitochondria. Ann Rev Biochem 76:723–749
Niedner A, Müller M, Moorthy BT et al (2013) The role of Loc1p in nuclear ASH1 mRNP assembly and reorganization in yeast. Proc Natl Acad Sci USA 110:5049–5058
Niessing D, Hüttelmaier S, Zenklusen D et al (2004) She2p is a novel RNA binding protein with a basic helical hairpin motif. Cell 119:491–502
Oeffinger M, Wei KE, Rogers R et al (2007) Comprehensive analysis of diverse ribonucleoprotein complexes. Nat Meth 4:951–956
Oleynikov Y, Singer RH (2003) Real-time visualization of ZBP1 association with beta-actin mRNA during transcription and localization. Curr Biol 13:199–207
Olivier C, Poirier G, Gendron P et al (2005) Identification of a conserved RNA motif essential for She2p recognition and mRNA localization to the yeast bud. Mol Cell Biol 25:4752–4766
Paquin N, Ménade M, Poirier G et al (2007) Local activation of yeast ASH1 mRNA translation through phosphorylation of Khd1p by the Casein Kinase Yck1p. Mol Cell 26:795–809
Peabody DS, Lim F (1996) Complementation of RNA binding site mutations in MS2 coat protein heterodimers. Nucleic Acids Res 24:2352–2359
Powrie EA, Zenklusen D, Singer RH (2010) A nucleoporin, Nup60p, affects the nuclear and cytoplasmic localization of ASH1 mRNA in *S. cerevisiae*. RNA 17:144
Prilusky J, Bibi E (2009) Studying membrane proteins through the eyes of the genetic code revealed a strong uracil bias in their coding mRNAs. Proc Natl Acad Sci USA 106:6662–6666
Reck-Peterson SL, Tyska MJ, Novick PJ, Mooseker MS (2001) The yeast class V myosins, Myo2p and Myo4p, are nonprocessive actin-based motors. J Cell Biol 153:1121–1126
Reid DW, Nicchitta CV (2012) Primary role for endoplasmic reticulum-bound ribosomes in cellular translation identified by ribosome profiling. J Biol Chem 287:5518–5527
Riedl J, Crevenna AH, Kessenbrock K et al (2008) Lifeact: a versatile marker to visualize F-actin. Nat Meth 5:605–607
Saint-Georges Y, Garcia M, Delaveau T et al (2008) Yeast mitochondrial biogenesis: a role for the PUF RNA-binding protein Puf3p in mRNA localization. PLoS ONE 3:e2293
Schmid M, Jaedicke A, Du T-G, Jansen R-P (2006) Coordination of endoplasmic reticulum and mRNA localization to the yeast bud. Curr Biol 16:1538–1543
Sellers JR, Veigel C (2006) Walking with myosin V. Curr Opin Cell Biol 18:68–73
Shen Z, Paquin N, Forget A, Chartrand P (2009) Nuclear shuttling of She2p couples ASH1 mRNA localization to its translational repression by recruiting Loc1p and Puf6p. Mol Biol Cell 20:2265–2275
Shen Z, St-Denis A, Chartrand P (2010) Cotranscriptional recruitment of She2p by RNA pol II elongation factor Spt4-Spt5/DSIF promotes mRNA localization to the yeast bud. Genes Dev 24:1914–1926
Shepard KA, Gerber AP, Jambhekar A et al (2003) Widespread cytoplasmic mRNA transport in yeast: identification of 22 bud-localized transcripts using DNA microarray analysis. Proc Natl Acad Sci USA 100:11429–11434
Sil A, Herskowitz I (1996) Identification of asymmetrically localized determinant, Ash1p, required for lineage-specific transcription of the yeast HO gene. Cell 84:711–722
Sladewski TE, Bookwalter CS, Hong M-S, Trybus KM (2013) Single-molecule reconstitution of mRNA transport by a class V myosin. Nat Struct Mol Biol 20:952–957
St Johnston D (2005) Moving messages: the intracellular localization of mRNAs. Nat Rev Mol Cell Biol 6:363–375
St Johnston D, Ahringer J (2010) Cell polarity in eggs and epithelia: parallels and diversity. Cell 141:757–774
Strathern JN, Herskowitz I (1979) Asymmetry and directionality in production of new cell types during clonal growth: the switching pattern of homothallic yeast. Cell 17:371–381
Takizawa PA, DeRisi JL, Wilhelm JE, Vale RD (2000) Plasma membrane compartmentalization in yeast by messenger RNA transport and a septin diffusion barrier. Science 290:341–344

Takizawa PA, Sil A, Swedlow JR et al (1997) Actin-dependent localization of an RNA encoding a cell-fate determinant in yeast. Nature 389:90–93

Urbinati CR, Gonsalvez GB, Aris JP, Long RM (2006) Loc1p is required for efficient assembly and nuclear export of the 60S ribosomal subunit. Mol Genet Genomics 276:369–377

Urbinati CR, Long RM (2011) Techniques for following the movement of single RNAs in living cells. WIREs RNA 2:601–609

Valverde R, Edwards L, Regan L (2008) Structure and function of KH domains. FEBS J 275:2712–2726

van den Bogaart G, Meinema AC, Krasnikov V et al (2009) Nuclear transport factor directs localization of protein synthesis during mitosis. Nat Cell Biol 11:350–356

Weis BL, Schleiff E, Zerges W (2013) Biochimica et Biophysica Acta. BBA—Mole Cell Res 1833:260–273

Wintersberger U, Kühne C, Karwan A (1995) Scp160p, a new yeast protein associated with the nuclear membrane and the endoplasmic reticulum, is necessary for maintenance of exact ploidy. Yeast 11:929–944

Yang H-C, Pon LA (2002) Actin cable dynamics in budding yeast. Proc Natl Acad Sci USA 99:751–756

Chapter 7
mRNA Degradation and Decay

Micheline Fromont-Racine and Cosmin Saveanu

Abstract Why is it important to understand mRNA degradation in a cell? First, RNA degradation has a clearing function and removes RNAs arising from transcription, splicing, export, or translation "accidents" to ensure robust gene expression (see Chap. 8). Second, while regulation of gene expression has a very important transcription component, mRNAs must be turned over rapidly for fast changes in transcriptome composition. Coordinated destabilization of an entire class of mRNAs can promote major physiological changes in a cell. Third, specific mechanisms of mRNA decay can serve to regulate gene expression through feedback control. Research on these topics has been frequently done first with yeasts and led to a better understanding of gene expression in eukaryotes. We start with an overview of the methods for measuring mRNA decay on a large scale with an emphasis on how technical issues affect the current picture of global mRNA decay in yeast. Next, we describe the importance of nuclear degradation in shaping the stable transcriptome. Once in the cytoplasm, mRNAs are exposed to translation and we provide an overview of the complexes and individual enzymes that ensure progressive deadenylation, mRNA decapping, and 5′ to 3′ or 3′ to 5′ exonucleolytic RNA degradation. Finally, how organelle transcripts are degraded in mitochondria is briefly exposed.

M. Fromont-Racine (✉) · C. Saveanu (✉)
Génétique des Interactions Macromoléculaires, Institut PASTEUR (UMR3525-CNRS),
25 rue du Docteur ROUX, 75015 Paris, France
e-mail: mfromont@pasteur.fr

C. Saveanu
e-mail: csaveanu@pasteur.fr

A. Sesma and T. von der Haar (eds.), *Fungal RNA Biology*,
DOI: 10.1007/978-3-319-05687-6_7,

Contents

Introduction

As a preliminary to describing how RNA is degraded in yeasts, we will first define what messenger RNA (mRNA) is. Probably one of the simplest definition is based on the mRNA potential to code for a functional protein, unlike noncoding RNAs such as tRNA or rRNA that are direct effectors in cellular machineries. To infer the protein coding potential and thus the inclusion of a transcript detected in a cell in the mRNA category, initial definition of open reading frames (ORFs) was done on the basis of length, starting at 100 codons. A powerful technique to go beyond such initial annotation is the comparison of genome sequence for many related species to identify conservation of predicted amino acid sequences (see, for example Dujon et al. 2004). The increase in the number of related genomic sequences leads to continuous improvements and reannotations of coding sequences in yeast (for an example of the evolution of *Saccharomyces cerevisiae* strain genomic sequence annotations over time, see Engel and Cherry 2013).

A pragmatic definition of an mRNA is based on its association with translating ribosomes. Such association can be tested by several methods, including

measurement of where an RNA sediments in a sucrose gradient (Arava et al. 2003), the extent of association with affinity purified ribosomes (Halbeisen et al. 2009), and the identification of ribosome-protected fragments (Ingolia et al. 2009). The final proof of coding potential and actual translation of an RNA is detection of the newly translated protein by mass spectrometry analysis (see, for example, Menschaert et al. 2013). This operational view of an mRNA also includes transcripts that do not produce functional polypeptides.

Studies of the factors affecting RNA decay and degradation allowed the discovery of novel concepts in transcription, nuclear export, and protein synthesis. Since the RNA degradation factors can have an action that is limited to one or another of the different cellular compartments, our description will follow the cellular organization of yeast cells. Nuclear, cytoplasmic, and mitochondrial processes will be described in separate sections.

Large-Scale Methods for mRNA Decay Measurements

What is the average half-life of each mRNA species in a yeast cell? In an era of deep RNA sequencing and refined methods for mRNA quantitation (see Chap. 15), the answer to this question could look trivial, but it is not. We will thus first take a look at the results of several methods used for mRNA decay measurements genome wide.

RNA Labeling with Modified Nucleotides

It would be highly convenient to be able to obtain statistics on the stability of individual RNA molecules in individual cells, and then calculate average values for an entire population of molecules and cells. Following the decay of single RNA molecules is possible but not yet at a large scale (Trcek et al. 2011). A traditional way for looking at molecules and their synthesis and decay uses radioactive labeling of living cells. A pulse of radioactively labeled compound that is incorporated in RNA can be used to follow the synthesis and, during a "chase" period, the disappearance of molecules over time. Such an approach has been instrumental, for example, in describing rRNA biogenesis intermediates in early studies done in the 70s (Trapman et al. 1975; Udem et al. 1971).

To establish the kinetics of synthesis and degradation of individual mRNA molecules in an organism, several criteria need to be met. First, labeling should not influence the process that is under study. This condition is very demanding since cell walls need to be removed or made permeable if labeled NTPs are to be incorporated. Second, labeled molecules should be easy to quantify and distinguish one from the other. Third, when cells grow fast, as many yeasts do, the dilution of the labeled molecule will impact the ability of the method to estimate long half-life

values. Dilution through active dividing cells during the assay needs to be taken into account (see Pelechano et al. 2010). Last, the localization of the degradation processes can affect the half-life estimations.

One of the first attempts to use a pulse labeling method to estimate indirectly mRNA half-life has been described as a genomic run-on strategy (García-Martínez et al. 2004). The method was based on previously described run-on experiments for transcription rate estimations (Hirayoshi and Lis 1999). Pulse labeling of newly synthesized RNA in yeast was done through a 5-min incubation with ^{33}P-UTP. Cells had been depleted of nucleotides and rendered permeable to UTP by detergent treatment (García-Martínez et al. 2004). Incorporation of UTP in specific RNA molecules was estimated by the use of nylon membranes on which PCR products encompassing annotated ORFs had been deposited. By a comparison between pulse labeling of RNA molecules and known steady-state levels of the corresponding mRNAs, estimates of half-life for thousands of yeast transcripts could be obtained. Enhancements and further corrections of the original data have been published a few years later (Pelechano et al. 2010). Nuclear degradation of some transcripts could affect the amount of newly synthesized RNA that reaches the cytoplasm (Gudipati et al. 2012). Thus, the stability of mature mRNA is likely to be underestimated by genomic run-on experiments. If a fraction of newly synthesized transcripts is degraded in the nucleus and another fraction in the cytoplasm, the kinetics of degradation could appear bimodal. Thus, real half-life values for cytoplasmic mRNAs could be longer than predicted from observed synthesis and steady-state levels.

Nonradioactive labeling of RNAs is a clever alternative to ^{33}P-UTP labeling. The incorporation of 4-thio-UTP in newly synthesized RNA (Fig. 7.1a and b) provides "hooks" that allow their isolation through biotinylation and affinity purification (Cleary et al. 2005; Dölken et al. 2008). The 4-thio-UTP precursor 4-thiouridine does not penetrate readily in yeast cells but its entry can be enhanced by the expression of the human nucleoside transporter hENT1 (Miller et al. 2011). Further analysis of the enriched RNA is performed by sequencing or DNA microarrays. 4-thiouridine-based methods are not devoid of problems. Two studies performed in two different laboratories (Munchel et al. 2011; Sun et al. 2012) have shown no correlation between the half-life estimates of yeast transcripts. Among several explanations for this lack of correlation, it is possible that the levels of 4-thiouridine used for pulse labeling RNAs were high enough to induce a stress response similar to the one observed in mammalian cells (Burger et al. 2013). Once the right conditions are set up, it is likely that 4-thiouridine labeling will turn out to be currently the most flexible method available for large-scale mRNA decay studies.

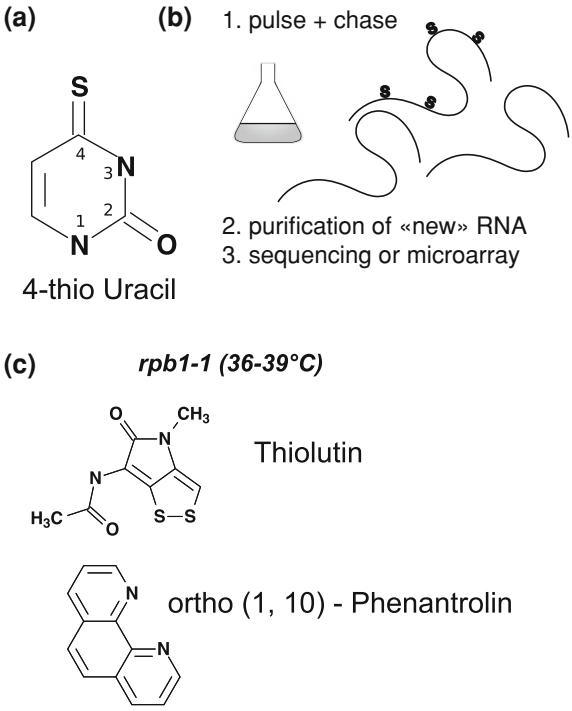


Fig. 7.1 Tools for large-scale analysis of mRNA turnover. **a** Structure of the modified U base that can be used in pulse-chase experiments; **b** Incorporation of 4-thioU in RNA by yeast cells during a pulse period is followed by the purification of newly synthesized RNA at different time points of "chase." The amounts of newly synthesized RNA are estimated through sequencing or DNA microarray analyses; **c** General blockers of mRNA transcription, including a temperature-sensitive mutant of an RNA polymerase subunit and two toxic chemicals used for turnover estimations

General Transcription Inhibition for Global mRNA Decay Tests

A straightforward and widely used method to assess mRNA stability is to follow what happens with an mRNA after transcription shutdown. Transcription inhibition is done either by using yeast strains with point mutations in an RNA polymerase II component (*rpb1-1*, Nonet et al. 1987) or by adding chemicals that are believed to specifically block RNA polymerase II, thiolutin (Jimenez et al. 1973) or ortho-phenantroline (Grigull et al. 2004) (Fig. 7.1c). An inherent problem of using a mutant strain defective for a major cellular pathway is that even at permissive temperature, transcription of the *rpb1-1* strain is reduced in comparison with a wild-type strain (Sun et al. 2012). It was even suggested that the measurements of half-life based on the study of such mutants are more likely to reflect changes in RNA stability due to heat shock. The very good correlation between the results of half-life estimates based on a temperature shift of an *rpb1-1* strain, or addition of thiolutin or phenantrolin (Grigull et al. 2004) indicate that these conditions affect transcripts stability by a common mechanism. Thiolutin treatment of the cells also inhibits mRNA degradation in a dose-dependent manner (Pelechano and Pérez-Ortín 2008), an effect that needs to be taken into account if estimates of half-life are done with this method.

For a more in-depth review on mRNA decay strategies, including a comparison with methods used for half-life estimations in other eukaryotes, see (Perez-Ortin et al. 2012). Half-life estimates of mRNA on a large scale need carefully crafted controls and independent methodological validation in different laboratories. Caution should be the rule when using any kind of half-life estimates for mRNA in yeast and probably in other species as well.

Linking mRNA Degradation with Global Transcription Changes

What knowledge can be gained from the study of mRNA half-life estimates and their changes in different environmental conditions? It was this type of studies that pioneered the idea that mRNA steady-state levels are a poor reflection of the relative role played by RNA degradation factors in cellular metabolism (Dori-Bachash et al. 2012; Shalem et al. 2008; Sun et al. 2012, 2013). Deleting genes for factors involved in mRNA degradation was accompanied by global transcription inhibition, leading to constant relative levels of mRNA in mutant strains. Contrary to the expected result, even if some classes of mRNA were affected differentially, the steady-state levels of mRNA were little affected by the absence of any given RNA degrading protein. A notable exception is the 5´ to 3´ exonuclease Xrn1, a major degradation factor in yeast cells (Sun et al. 2013). Why Xrn1 has an effect that is different from other degradation factors on the relationship between transcription inhibition and mRNA decay remains unknown.

An explanation for how RNA degradation and synthesis could be linked comes from the observation that mRNAs for transcription repressors, like Nrg1, increase in degradation deficient strains (Sun et al. 2013). As a consequence, the steady-state level of the protein is likely to rapidly increase, which leads to general repression of transcription. *NRG1* mRNA is known to be under tight degradation control since its levels increase in mutants depleted for the major deadenylase Ccr4 (Lo et al. 2012). *NRG1* transcript levels also increase upon translation inhibition with cycloheximide (Sun et al. 2012). Thus, deletion of genes for RNA degradation factors could mimic physiological situations that modulate translation. The tight coupling between translation, translation regulation, and mRNA degradation of transcription factors could ensure robust responses of yeast cells to environmental changes and stress. Transcription rate of thousands of genes can be affected by changes in translation or degradation of a few key mRNAs. As a consequence of yet unknown secondary effects, steady-state levels of transcripts are a poor predictor of the direct effect of a degradation factor or complex perturbation on RNA. Tests of direct binding of degradation factors to RNA, as explained in the following section, are likely to be more effective in finding physiological roles for proteins involved in RNA stability.

mRNA Stability, Associated Proteins, and the RNA Operon Concept

The first experimental observations that identified a potential role of RNA-binding proteins as global gene expression regulators were made in human cells (Tenenbaum et al. 2000). Based on these data and on previous knowledge on the properties of RNA-binding proteins and especially their ability to change mRNA turnover, Keene and Tenenbaum proposed the RNA operon concept (Keene and Tenenbaum 2002; reviewed in Keene 2007), which states that a given protein can be part of a large number of mRNPs and can play key roles in affecting the turnover of a large number of mRNAs that code for functionally related factors. While these ideas started from work done with mammalian cells, an experimental confirmation of this hypothesis came later from yeast experiments performed in Brown's laboratory. The development of affinity-based purification of RNAs in association with specific proteins coupled with microarrays allowed some of the first large-scale estimations of the complex picture of the mRNP world in yeast (Gerber et al. 2004; Hogan et al. 2008).

The best-known cases of RNA operons involve proteins of the Puf family. Described initially in *Drosophila melanogaster* and *Caenorhabditis elegans*, there are six Puf proteins in *S. cerevisiae* (reviewed in Quenault et al. 2011; Wickens et al. 2002). The Puf proteins share eight repeats of the Puf motif that forms an arc-like structure allowing interaction with both RNA and other protein cofactors. Subtle amino acid changes allow specific interaction with slightly different RNA motifs for each Puf protein (Qiu et al. 2012). While some Puf proteins overlap in their specificity of binding to 3´ UTR regions of yeast transcripts (Puf1, Puf2 for membrane proteins mRNA), others are highly specific to particular mRNA classes (Gerber et al. 2004). Puf4 prefers mRNAs for ribosome biogenesis factors while Puf5 is mostly devoted to chromatin modification factor mRNAs.

Puf3 binds to many transcripts coding for proteins that are imported co-translationally into mitochondria. Their 5´ regions of translated polypeptides contain a sequence of 60 amino acids coding for a mitochondrial targeting sequence, which contributes to mRNA targeting to mitochondria. The Puf3 protein is required for the specific intracellular localization of one of two classes of mitochondria targeted transcripts (Saint-Georges et al. 2008). While more than 200 transcripts are localized to mitochondria in a Puf3-dependent manner and contain consensus binding sequences for Puf3 in their 3´ UTR region, other mRNAs that are translated in the vicinity of the mitochondria are not affected by Puf3. Puf3 binding to mitochondrial transcripts contributes not only to their localization but also stimulates deadenylation and modulates the stability of the bound mRNA (Foat et al. 2005; Jackson et al. 2004; Olivas and Parker 2000). An MFA2-COX17 chimeric transcript bearing the 3´ UTR of *COX17*, coding for a mitochondrial protein, thus sees its half-life increased from 3 to 10 min in a *puf3*Δ strain. Whether Puf proteins directly activate deadenylation by interacting with Pop2 and recruiting the Pop2-Ccr4-Not complex (Goldstrohm et al. 2006, 2007), modify

decapping, or inhibit translation, the net effect of these mechanisms is a change in the stability of the class of mRNA to which these proteins bind.

The RNA operon concept is not limited to Puf proteins. Many other abundant RNA-binding proteins potentially affect the stability of classes of mRNA. The formation of RNA protein complexes in the nucleus and the factors that affect mRNA export can influence whole classes of transcripts. Hrp1 (Nab4), an RNA binding shuttling protein that is required for mRNA export was found to strongly bind to transcripts encoding proteins involved in amino acid metabolism (Kim Guisbert et al. 2005). Changes of a single protein, like Hrp1, can thus affect the availability of a large number of mRNAs for cytoplasmic translation and their stability.

Analysis of mRNP composition by affinity purification of RNA bound to RNA-binding proteins is not without technical biases and can lead to erroneous results. The use of porous agarose-based affinity matrices leads to a heavily shifted representation of various classes of transcripts. Very large mRNPs are excluded from binding to chromatographic beads, unless those beads are small and compact and only allow affinity binding on their surface (Halbeisen et al. 2009). Fortunately, a more fine-grained image of mRNP composition became possible through the development of cross-linking and sequencing methods, best illustrated in yeast studies by the CRAC technique (Granneman et al. 2009), a variation of the widely used CLIP approach (Ule et al. 2003) (more details in Chap. 14). Recent data on sets of RNA-binding proteins and their RNA targets have further enriched our knowledge of mRNP composition at different stages of mRNA maturation or destruction (Klass et al. 2013; Mitchell et al. 2013; Tuck and Tollervey 2013). Various methods have been used to analyze either the protein composition of poly (A)-bound material (Garland et al. 2013; Mitchell et al. 2013) or the RNA composition of protein-associated complexes stabilized through UV cross-linking (Tuck and Tollervey 2013). These recent data are rich in information and are likely to vastly expand the knowledge on mRNA–protein interactions. Correlations between the various large-scale datasets are expected to bring more detail and mechanistic insights into which RNAs bind which proteins and in what physiological context these interactions are relevant.

Nuclear Degradation of Nascent mRNA

The mRNA cap and poly(A) tail are the most important determinants of mRNA stability. Uncapped nascent transcripts, intermediates in mRNA formation, and even mRNA that are not efficiently exported from the nucleus are the subject of degradation mechanisms that are confined to the nucleus. In addition to nuclear quality control of RNA, co-transcriptional recruitment of RNA-binding proteins can affect the cytoplasmic stability of the corresponding mRNA (Bregman et al. 2011; Trcek et al. 2011).

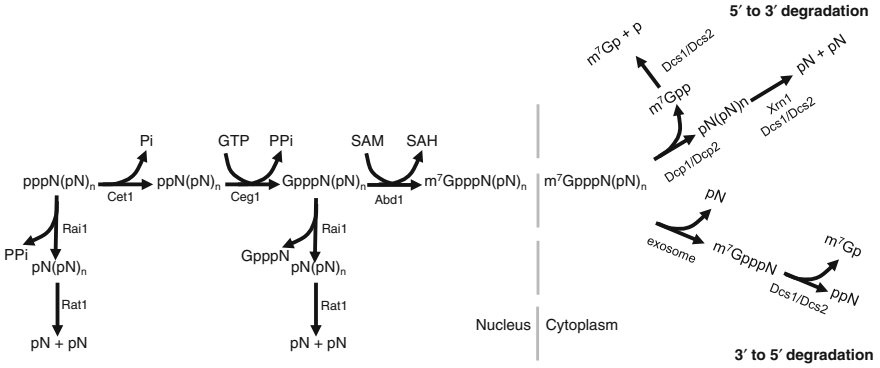


Fig. 7.2 Formation and degradation of capped mRNA and *S. cerevisiae* enzymes involved in the process

mRNA Capping and Nuclear Quality-Control of the 5′ End

mRNA capping is one of the first co-transcriptional modifications that occurs when nascent pre-mRNA reaches 22–25 nucleotides in length (in human cells, Moteki and Price 2002). The capping complex containing RNA 5′ triphosphase and guanyltransferase activities was first purified from vaccinia virus (Ensinger et al. 1975). Forty years later, mechanistic details about three successive steps are available (Fig. 7.2). Cet1, an RNA 5′ triphosphatase, hydrolyzes the gamma phosphate from the 5′ end of the nascent transcript. This step is followed by a transfer of GMP to the diphosphate 5′ end by Ceg1, an RNA guanyltransferase. Finally, Abd1, an RNA N^7 Guanine methyltransferase transfers a methyl group on the guanine base at the N^7 position (reviewed in Shatkin and Manley 2000). The Cet1/Ceg1 heterodimer interacts with the Ser5 phosphorylated form of the CTD of RNA polymerase II. Based on structural studies of the capping complex (Gu et al. 2010), it was proposed that Cet1/Ceg1 are recruited by the Ser5-phosphorylated CTD near the transcription start site. Recruitment of Abd1 is maximal at about 100 nt downstream transcription start and is almost concomitant with the binding of the cap-binding complex (CBC). Both Abd1 and CBC are required for the recruitment of the kinases Ctk1 and Bur5, which release the capping enzymes and promote RNA polymerase II elongation (Lidschreiber et al. 2013). If the capping process is erroneous, the resulting nascent transcripts are degraded from the 5′ end in a process involving the exonuclease Rat1 and the associated factor Rai1 (Jiao et al. 2010).

RAT1 was identified through a genetic screen for factors affecting RNA export from the nucleus (Amberg et al. 1992). The protein is a nuclear 5′ to 3′ exoribonuclease, which is similar in sequence with the cytoplasmic exonuclease Xrn1 and is partially functionally redundant in yeast (Johnson 1997). Rat1 actively degrades RNA with a monophosphorylated 5′-end. Such an end can be generated by the decapping activity of Rai1, a Rat1 cofactor, that can remove the dinucleotide cap, especially if it is not methylated (Jiao et al. 2010) (Fig. 7.2). Completion

of the capping reaction is affected under amino acid or glucose starvation, physiological conditions that might require rapid changes in mRNA transcription, export and degradation. Biochemical data also indicate that Rai1 stabilizes and stimulates Rat1 activity, which alone is relatively unstable (Xue et al. 2000), probably through the formation of heterodimers (Xiang et al. 2009). In addition to the involvement of Rat1 in the 5′ to 3′ degradation of partially capped or uncapped nascent transcripts, the enzyme also participates in transcription termination. Active degradation of RNA fragments synthesized by an RNA polymerase II after the cleavage step could serve in the removal of the polymerase from the DNA template ("torpedo" termination model Luo et al. 2006; more details in Chap. 1). The importance of the catalytic activity of Rat1 in this context is not clear (Pearson and Moore 2013).

Nuclear Retention and Transcript Degradation

Recognition and degradation of aberrant or intermediate mRNA forms that fail to mature properly is essential for the control of gene expression. Both nuclear and cytoplasmic degradation mechanisms are required for quality control of RNA. The balance between degradation into the nucleus or cytoplasm depends on export or retention of RNAs. Unspliced pre-mRNAs that are exported to the cytoplasm are degraded by nonsense-mediated mRNA decay (NMD) (see Chap. 8), but aberrant pre-mRNAs retained in the nucleus are efficiently degraded. One of the first observations of nuclear retention of unspliced pre-mRNA was done with beta-galactosidase reporters designed to distinguish exported unspliced pre-mRNA from the spliced form (Legrain and Rosbash 1989). The involvement of the exosome in the degradation of pre-mRNA in the nucleus and functional links between 3′ and 5′ intranuclear degradation of transcripts and splicing were demonstrated later on endogenous RNAs (Bousquet-Antonelli et al. 2000). In *Cryptococcus neoformans*, an organism in which all mRNAs are generated from spliced transcripts, splicing seems an absolute requirement for mRNA export. Transcripts without introns are retained and degraded in the nucleus in a process dependent on *C. neoformans* Pab2 (Goebels et al. 2013), a protein related to *Schizosaccharomyces pombe* Pab2 (Lemieux and Bachand 2009) and to *S. cerevisiae* Sgn1 (of unknown function).

If nuclear degradation is not effective, the yeast cells are able to degrade pre-mRNAs in the cytoplasm through NMD. Studies of cells with defective or absent *RRP6*, a nuclear 3′ to 5′ exonuclease and *NAM7* (*UPF1*) a major NMD factor, indicate that nuclear exosome and cytoplasmic NMD pathways could act sequentially. Blocking RNA export in a *mex67* mutant strain leads to an accumulation of unspliced mRNA in the absence of *RRP6* (Sayani and Chanfreau 2012). Some pre-mRNAs are degraded predominantly by nuclear mechanisms while others are exported and degraded in the cytoplasm.

Nuclear Degradation from the 3′ End: The Exosome

The nuclear exosome is a conserved complex of proteins mentioned in the previous section as a key player in the 3′ to 5′ exonucleolytic degradation of RNA synthesized by all the eukaryotic RNA polymerases (tRNA, rRNA, nascent mRNA). The core exosome is present both in the nucleus and the cytoplasm and consists of nine subunits (Rrp4, Rrp40, Ski6/Rrp41, Rrp42, Rrp43, Rrp45, Rrp46, Csl4, and Mtr3) with which Dis3/Rrp44 is strongly associated (Mitchell et al. 1997). The nuclear and cytoplasmic forms of the exosome differ in their binding partners and posttranslational modifications of the components. Phosphorylation of Csl4 at Ser94 was found to be specific for the nuclear version of the exosome, associated with Rrp6, and is less present in the cytoplasmic version, associated with Ski7 (Synowsky et al. 2009). While the core exosome components are related to RNAses, only Dis3/Rrp44 is an active enzyme (Dziembowski et al. 2007; Schneider et al. 2007) through two catalytic domains: one that belongs to the RNAse II 3′ to 5′ exonuclease family and a second, endonucleolytic PIN domain (Lebreton et al. 2008; Schaeffer et al. 2009; Schneider et al. 2009). In addition to Dis3, Rrp6 is a strictly nuclear exonuclease of the 5′ to 3′ DEDD family of RNases (reviewed in Arraiano et al. 2013) associated with the nuclear core exosome.

The nine subunits of the core exosome are essential for yeast viability and form a ring structure (Liu et al. 2006). Mutations occluding the ring channel inhibit endo- and exonucleolytic activities suggesting that the core exosome modulates Rrp6 and Dis3 activity (Wasmuth and Lima 2012). Unwound RNA substrates enter into the internal chamber formed by the nine inactive subunits organized in a barrel-like structure (Makino et al. 2013a) and progress through the chamber up to the catalytic site of Dis3, which interacts with the bottom of the core exosome. Rrp6 is associated with the side in proximity of the mRNA entry point in the barrel-like structure and its catalytic site is exposed at the surface of the enzyme. The unwinding activity for the entry of mRNA into the internal chamber of the exosome is probably provided by Mtr4, a helicase of the DExH family that is similar to Ski2, a cytoplasmic RNA helicase associated with the exosome.

Mpp6 and Rrp47

Deletion of either *MPP6* or *RRP47/LRP1* has a negative effect on the growth of strains depleted for Rrp6 (Milligan et al. 2008 and CS, unpublished). Both proteins are also physically associated with Rrp6 and their function is partially redundant since deletion of the corresponding genes leads to a growth defect (Milligan et al. 2008). Rrp47 concomitantly interacts with Rrp6 and RNA through its N-terminal and C-terminal domains, respectively (Costello et al. 2011; Stead et al. 2007). In contrast to Mpp6, which recognizes pyrimidine-rich sequences, Rrp47 binds to

structured RNA molecules suggesting that Rrp47 could promote the binding of Rrp6 to substrates (Butler and Mitchell 2011 for review; Garland et al. 2013).

Mpp6 function and role in RNA degradation remain elusive. Deletion of the corresponding gene leads to growth defects specifically in association with concomitant deletion of genes for components of the TRAMP complex. Its association with Rrp6 in the nucleus as well as the effects of its deletion, revealed that Mpp6 is involved in surveillance and degradation of nuclear pre-mRNAs and pre-rRNA (Milligan et al. 2008). It was proposed that Mpp6 could also promote the activity of Dis3 and be involved in the functional coupling between Rrp6 and the TRAMP complexes (for review, Butler and Mitchell 2011).

TRAMP Complexes

The nuclear exosome is helped by additional factors for degradation of its RNA substrates. A major functional nuclear co-factor of the exosome that adds poly(A) tails to nuclear RNA is the TRAMP complex. TRAMP complexes are formed of a poly(A) polymerase (Trf4 or Trf5), an RNA helicase of the DexH family (Mtr4) and of a Zn-knuckle RNA-binding protein (Air1 or Air2) (LaCava et al. 2005; Vanácová et al. 2005; Wyers et al. 2005). At least two TRAMP complexes with different substrate specificity have been described. A major difference between these complexes is the Trf component, with Trf5-TRAMP most likely located to the nucleolus (Wery et al. 2009) and Trf4-TRAMP located into the nucleoplasm. A comparative transcriptome analysis of *trf4*Δ or *trf5*Δ mutants indicates that the two proteins affect the expression of distinct sets of genes (San Paolo et al. 2009). In addition, Air1 is mostly present in the Trf5-TRAMP complex while Air2 is mainly present in Trf4-TRAMP (reviewed in Houseley et al. 2006). These differences correlate well with recent RNA sequencing data obtained with *air1*Δ and *air2*Δ mutants, which revealed, as expected, a different global effect of each mutant on transcripts levels (Schmidt et al. 2012). A TRAMP complex would bind the RNA targets through its RNA-binding subunit Air1 or Air2 and add a poly(A) tail through its Trf4/5 polymerase subunit (Holub et al. 2012). The addition of poly(A) tails allows better access of the target to the nuclear exosome. The TRAMP complex is also able to enhance RNA degradation by Rrp6 independently of the presence of exosome in vitro (Callahan and Butler 2010).

The Nrd1/Nab3/Sen1 Complex (NNS)

Deciding whether an mRNA precursor will be stable until export to the cytoplasm or not was shown to depend in most instances on the way transcription by RNA polymerase II ends. Early transcription termination can occur by a pathway that is linked with nuclear processing or degradation of the corresponding RNA. Acting upstream of the nuclear exosome and TRAMP complexes, this pathway depends on a transcription termination complex that marks the corresponding RNAs as

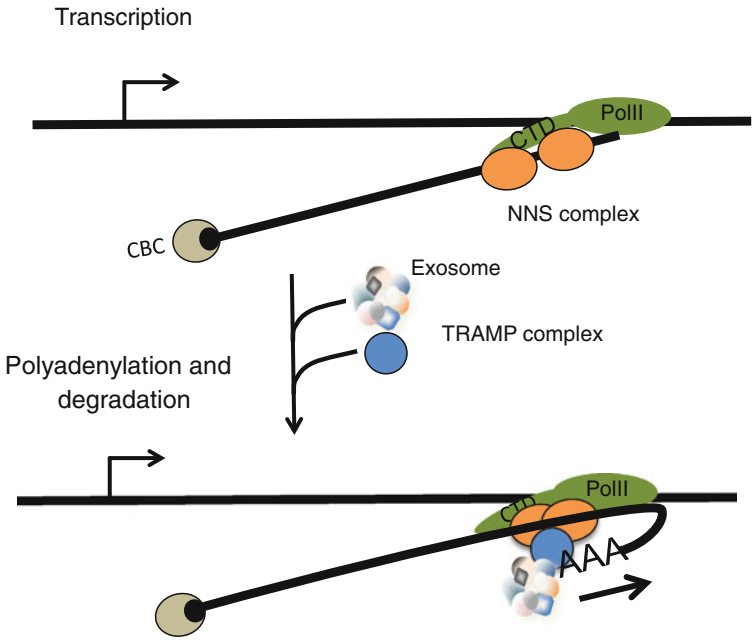


Fig. 7.3 Nrd1-Nab3-Sen1 play an important role in an alternative transcription termination pathway that is linked with poly(A) addition by the TRAMP complex and degradation by the nuclear exosome. Such termination is most efficient on short transcripts

exosome targets (Fig. 7.3). The NNS complex is formed of Nrd1 and Nab3, two RNA-binding proteins, which preferentially recognize short RNA motifs (such as GUAA and UCUUG, Porrua et al. 2012) and Sen1, an RNA helicase. In contrast to the cleavage and polyadenylation complex that is required for transcription termination of most mRNAs (see Chap. 3), the NNS complex is involved in transcriptional termination of cryptic unstable noncoding RNA (CUT) and of stable noncoding RNA like snoRNA and snRNA (reviewed in Jensen et al. 2013). Nrd1/Nab3-binding sites are underrepresented in mRNAs, which are thus less sensitive to the NNS-TRAMP-exosome-dependent termination degradation pathway.

Nuclear mRNA Degradation and Regulation

NNS termination can play an important role in regulating mRNA levels. *RPL9B,* a gene coding for a ribosomal protein of the large 60S subunit has a choice between the two modes of transcriptional termination that depends on the level of nuclear Rpl9 protein. The protein binds a stem-loop located in the 3´ UTR of the primary transcript of *RPL9B* and inhibits normal transcription termination. Alternative

termination through NNS leads to efficient nuclear degradation of the transcript and effectively regulates mRNA levels for *RPL9B* (Gudipati et al. 2012).

NRD1 expression is regulated by premature transcription termination through the interaction of the Nrd1 protein with its own mRNA (Arigo et al. 2006). High-throughput analyses through crosslinking, protein purification, and RNA sequencing revealed that Nrd1-Nab3 complexes are recruited during transcription of a large number of mRNAs, suggesting that the NNS complexes could be widely involved in mRNA downregulation (Schulz et al. 2013; Wlotzka et al. 2011).

Transcript retention or export plays an important role in deciding which degradation pathway a given molecule will take. Polyadenylation of transcripts plays an important role in nuclear export. Co-transcriptional recruitment of Nab2, a nuclear poly(A)-binding protein (Anderson et al. 1993) is one of the molecular events that shape the export-competent mRNA protein complexes (Green et al. 2002; Hector et al. 2002). Nab2 interaction with Mex67 and Yra1, mRNA export factors, could ensure recruitment of the Mex67-Mtr2 complex (Iglesias et al. 2010), and further interactions with the nuclear pore proteins. In addition to a general role in mRNA export, Nab2 regulates its own expression levels by acting at the level of the 3′ end formation of its own mRNA. This process depends on a genome-encoded repeat of 26 adenosines found downstream the stop codon for *NAB2* (Roth et al. 2005). The autoregulation of *NAB2* depends on the encoded poly (A) sequence and also requires the nuclear exosome and TRAMP complexes (Roth et al. 2009). *NAB2* is thus an example of autoregulation mechanism that uses nuclear degradation of a transcript to reduce gene expression.

Cytoplasmic mRNA Degradation

Cytoplasmic mRNA decay occurs mainly from both the 5′ and the 3′ end, with little or no endonucleolysis occurring in yeast. Whatever the pathway, deadenylation is considered to be the first step in mRNA turnover. In 1975, Darnell's laboratory observed a correlation between poly(A) tail shortening and mRNA turnover in HeLa cells (Sheiness et al. 1975). Using stable and especially the *MFA2* unstable mRNA, it was later established that deadenylation is the first step required for mRNA decay in *S. cerevisiae* (Muhlrad et al. 1994). Two complexes, Pan2/Pan3 and Ccr4/Not, are involved in the deadenylation process. Once deadenylated, mRNA is predominantly degraded by the 5′ to 3′ degradation pathway, which is initiated by the removal of the cap structure by the Dcp2/Dcp1 complex with the help of enhancers of decapping. After decapping, the 5′ end of mRNA is accessible to the cytoplasmic exonuclease Xrn1, related to the nuclear exonuclease Rat1. Cytoplasmic mRNA degradation and translation are tightly related through translational repressors that also act as activators of decapping. After deadenylation, mRNA can also be degraded by the 3′ to 5′ pathway. This degradation is done by the cytoplasmic exosome with help from specific cytoplasmic cofactors,

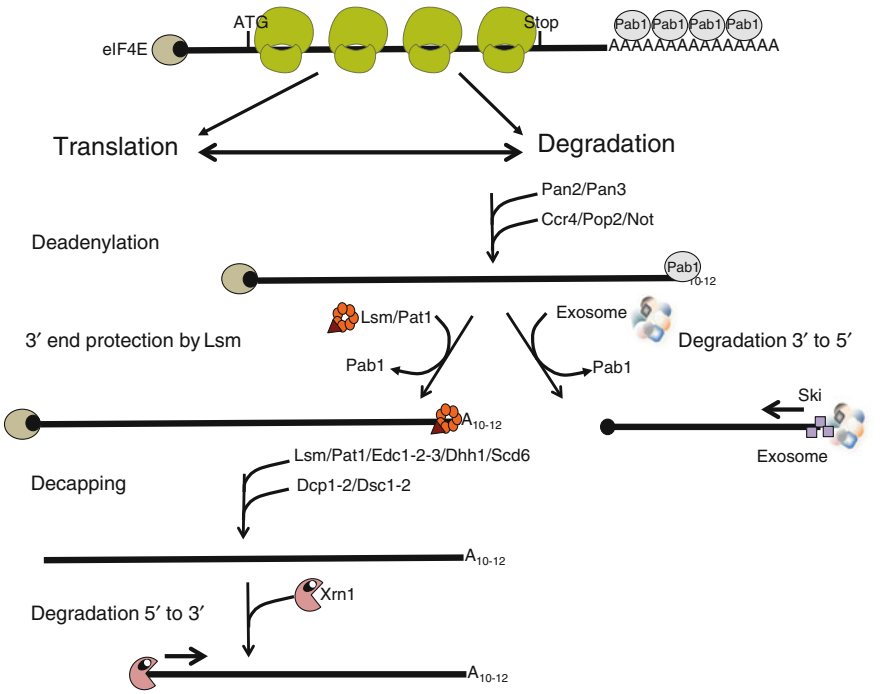


Fig. 7.4 Pathways for the cytoplasmic degradation of mRNA. Known functions and substrates of the different decay factors are discussed in the text

the Ski complex and Ski7. An overview of the cytoplasmic degradation pathways is presented in Fig. 7.4.

Cytoplasmic mRNA Deadenylation

Poly(A)-Binding Proteins Pab1 and Pub1

Pab1, for poly(A)-binding protein, was first isolated and its gene identified by Sachs and Kornberg (Sachs et al. 1986). This protein binds specifically poly(A) through four RNA recognition motifs (RRM); however the first two RRMs are sufficient for most in vivo functions of the protein. A minimum of 12 adenosines is required for the association and multiple associations define a coverage unit of 27 adenosines (Sachs et al. 1987). One of the best characterized roles of Pab1, which decorates the 3′ end of transcripts like beads on a string (Baer and Kornberg 1980, 1983), is to assist translation initiation through specific interactions with the translation initiation factor eIF4G (Tif4631/Tif4632 in yeast). eIF4G binds the cap-binding protein eIF4E (Cdc33 in *S. cerevisiae*). These interactions lead to

the formation of a "closed loop" structure, which was proposed to promote translation initiation and ribosome recycling (Amrani et al. 2008; see Chap. 5). Depletion of Pab1 leads to an inhibition of poly(A) tail shortening and has strong functional interactions with factors involved in 60S ribosomal subunit biogenesis (Sachs and Davis 1989). Thus, paradoxically, Pab1 both protects the polyA from degradation and contributes to the controlled 3′ end trimming, presumably by recruiting the Pan2/3 complex (Mangus et al. 2004b) and by inhibiting the deadenylase activity of Ccr4/Not (Tucker et al. 2002).

In addition to Pab1 and Nab2, which are mainly cytoplasmic and nuclear, respectively, Pub1 is a third polyA-binding protein present in both cellular compartments but mainly located in the cytoplasm (Matunis et al. 1993). Pub1 also recognizes poly(U) stretches and interacts with 5′ and 3′ UTR regions in several mRNAs. In contrast to Pab1, Pub1 is not associated with translationally active mRNAs (Anderson et al. 1993b). Pub1 is abundant but not essential for growth of cells under laboratory conditions. The protein is involved in mRNA stability and translational control under environmental stress and colocalizes with Pab1, eIF4G1, and eIF4G2, *inter alia*, in granules in glucose-deprivation stress and is required for the formation of these granules (Buchan et al. 2008). Of the three RRM motifs found in Pub1, at least one is required for an interaction with eIF4G. It was thus proposed that Pub1 could act cooperatively with Pab1 to simultaneously interact with eIF4G (Santiveri et al. 2011).

The Pan2/Pan3 Deadenylase

PAN stands for poly(A) nuclease and corresponds to a cytoplasmic enzymatic activity that shortens poly(A) tails of yeast transcripts only in the presence of the poly(A)-binding protein Pab1. The identification of the PAN complex components was the result of a purification procedure that enriched a Pab1-dependent deadenylase activity (Sachs and Deardorff 1992). Initially, a co-purifying protein, called Pan1, was thought to be the enzyme required for the observed Mg^{2+}-dependent exonuclease activity. However, an increase in the amounts of purified complex and further analyses around the role of Pan1 in the catalytic activity of the purified fraction have shown that the isolated deadenylase consisted of two subunits: Pan2, the enzyme (Boeck et al. 1996) and Pan3, a cofactor (Brown et al. 1996). While the catalytic center belongs to Pan2, Pan3 is required to an equal extent for the deadenylase activity of the complex. The Sachs group, involved in the initial identification of the Pan proteins demonstrated later that the Pan2/Pan3 complex plays a role in the modulation of poly(A) tail length distribution in yeast (Brown and Sachs 1998).

Pan2 is a 3′ to 5′ exonuclease of the DEDD family (ribonucleases reviewed in Arraiano et al. 2013) that slowly removes 5′ AMP from the 3′ end of a poly(A) tail only in the presence of Pab1 but will not proceed to lengths inferior to 20 nucleotides in vitro (Lowell et al. 1992). The enzyme generates in vitro the entire range

of oligo(A) intermediates, showing a distributive mode of action. The in vivo role of Pan2/Pan3 in deadenylation of mRNA remains poorly characterized.

The Ccr4/Not Deadenylase

Deadenylation of mRNA in the cytoplasm requires the Ccr4/Not complex, also called the major mRNA deadenylase complex. Ccr4/Not complex is formed of nine subunits (Not1, Not2, Not3, Not4, Not5, Ccr4, Pop2, Caf40, and Caf130). Not1 is the scaffold protein on which various partners bind: the C-terminal part recruits Not2, Not3, Not4, and Not5 (the Not module) and the N-terminal part recruits Ccr4 and Pop2 (the nuclease module) (Basquin et al. 2012). Ccr4 bears the main catalytic activity (Tucker et al. 2002) while Pop2 contains a RNAse D domain and has an exonuclease activity in vitro (Daugeron et al. 2001). To what extent Ccr4 and Pop2 cooperate in the deadenylation process remains unclear. The deletion of the other genes of the complex only weakly slows down deadenylation of a model substrate. It was proposed that the Not module could adapt the deadenylase complex to mRNA according to the cellular context. To gain a better understanding of the role of each component of the Not module, genome wide analyses using deletion mutants suggested that the Ccr4/Not complex is involved in a number of other cellular functions (for review, see Collart and Panasenko 2012).

The Pan2/3 and Ccr4/Not complexes in association with Pab1 are involved in 3′ end deadenylation. It is believed that Pan2/3 acts first to shorten the poly(A) of newly synthesized mRNA. Next, it is the Ccr4 complex that deadenylates mRNAs until the tail reaches a length of 10 to 12 residues, and can be bound by the Lsm complex and Pat1 that trigger mRNA decapping (for a review, Parker 2012).

Cytoplasmic Degradation from the 3′ End: The Exosome with Ski

The cytoplasmic exosome contains the same factors as the nuclear version but associates with the cytoplasmic Ski complex and Ski7, instead of nuclear Rrp6 or the TRAMP complex. The Ski complex is formed of three proteins and was initially identified for its role in the degradation of viral RNA in yeast. Yeasts deficient for *SKI* genes have a super-killer phenotype because a virally encoded toxin is expressed at higher levels and kills neighboring susceptible yeasts (Toh-E et al. 1978; Widner and Wickner 1993). While Ski2 is an RNA helicase related to nuclear Mtr4, Ski3 and Ski8 contain structural motifs that allow protein–protein interactions; tetratricopeptide repeats (TPR) for Ski3; and WD40 repeats for Ski8. Native mass spectrometry experiments showed that the Ski complex is a hetero-tetramer composed of two molecules of Ski8 and one copy of Ski2 and Ski3 (Synowsky and Heck 2008). Recently, the crystal structure of the complex

revealed its organization in which Ski3 plays a scaffold role. RNase protection assays on RNA in presence of exosomes with or without Ski7 and the Ski complex revealed RNA fragments of 43–44 or 31–33 nt, respectively, indicating that fragments of RNA are protected in the exosome channel (Halbach et al. 2013). Interestingly, the large size of the protected fragments was compatible with the sum of the size of the channel of the Ski complex and the exosome channel suggesting that the two complexes are stacked in close proximity. Therefore, the structural organization of the exosome and Ski complex bears similarities with the proteasome structure. The Ski complex would play the role of the regulatory 19S proteasome subunit and the core exosome would be structurally similar to the 20S proteasome (Makino et al. 2013b).

The role of Ski7, which strongly associates with the cytoplasmic exosome remains elusive. No structural data exists on Ski7 but an association of the protein with the Ski complex (Araki et al. 2001; Wang et al. 2005) and with the Csl4 subunit of the exosome (Schaeffer et al. 2009) were shown. Ski7 homologs cannot be found outside yeasts in contrast to the components of the Ski complex and exosome, which are highly conserved in eukaryotes. It has been proposed that yeast Ski7 resulted from divergent evolution of a duplicated ancestral Ski7/Hbs1 gene (Marshall et al. 2013). Hbs1 is a conserved protein related to translation termination factors and involved in solving the problem of ribosomes stalled on mRNA. Subfunctionalization of Hbs1 and Ski7 was potentially influenced by the requirement for maintaining yeast viral systems that have no equivalent in other eukaryotes (see Drinnenberg et al. 2011).

In contrast to *S. cerevisiae*, where only one version of *DIS3* exists, *S. pombe* cells contain a related enzyme called *DIS3L2*. This exonuclease is cytoplasmic, shows preference for poly(U)-tailed transcripts and does not require the rest of the exosome subunits for its activity (Malecki et al. 2013). Poly-urydilated substrates in *S. pombe* could result from the action of Cid1, an enzyme related to Trf4 and Trf5 (Wang et al. 2000). Cid1 is cytoplasmic and shows both a poly(A) polymerase activity and a significant poly(U) polymerase activity on a model RNA substrate in vitro (Read et al. 2002). Uridylation of substrates does not require prior deadenylation and facilitates the binding of the Pat1/Lsm complex for decapping and mRNA degradation (Rissland and Norbury 2009). The specific role of Dis3l2 and the number and importance of substrates that are poly-uridylated for degradation in *S. pombe* remains to be investigated.

Cytoplasmic Degradation from the 5′ End

The Cytoplasmic Decapping Enzyme (Dcp1-Dcp2)

In 1976, J. Warner's laboratory published the observation that yeast mRNA contains a modified 5′ end with either m7G(5′)pppAp or m7G(5′)pppGp (Sripati et al. 1976). A correlation between the 5′ end structure and mRNA stability was

later described by experiments with viral mRNA injected into *X. laevis* oocytes or incubated with cellular extracts from mouse cells (Furuichi et al. 1977). An important conclusion of this study was that capped mRNAs were not degraded from 5´ to 3´ and their stability was independent of translation. In 1980, a decapping activity was detected from a high-concentration salt wash ribosomal fraction by Audrey Stevens who purified a decapping enzyme from *S.cerevisiae* (Stevens 1980, 1988). A decade later, the first decapping enzyme gene, *DCP1* was identified by Parker´s laboratory (Beelman et al. 1996). Genetic screens using a *dcp1-2 ski8*Δ strain, allowed the isolation of a second decapping gene, *DCP2* (Dunckley and Parker 1999). Human Dcp2 was shown to bear the decapping catalytic activity (van Dijk et al. 2002) and in yeast it was also shown that Dcp1 plays the role of an auxilliary, albeit important, factor for the Dcp1/2 heterodimer (Steiger et al. 2003). The binding of Dcp1 to the N-terminal domain of Dcp2 was later shown to promote the catalytic activity without affecting the binding of RNA to the C-terminal Nudix domain of Dcp2 (Deshmukh et al. 2008).

Dcp2 is not the only decapping enzyme in yeast. Dxo1, sharing a weak homology with Rai1 was identified as a novel decapping enzyme, which also has a 5´ to 3´ exonuclease activity (Chang et al. 2012). Global GFP-fusion protein localization indicates that Dxo1 is mainly present in the cytoplasm. However, additional studies are required to determine whether Dxo1 is strictly cytoplasmic or could be also involved in nuclear decapping processes.

Activators of Decapping: Edc1, Edc2, and Edc3

Two regulators or "enhancers" of decapping, Edc1 and Edc2, were isolated from a genetic screen looking for genes whose overexpression could restore the viability of a *dcp1-2 ski8*Δ strain (Dunckley et al. 2001). These proteins bind RNA and directly interact with Dcp1 via their proline-rich regions to stimulate the activity of the decapping enzyme (Schwartz et al. 2003; Borja et al. 2011).

Edc3 was originally selected in two-hybrid screens as physical partner of Lsm proteins, Dcp2 and Xrn1, suggesting that the protein could play a role in mRNA decay (Fromont-Racine et al. 2000). Unlike Edc1 and Edc2, Edc3 is a conserved protein in most eukaryotes and its sequence contains an Sm-domain (Albrecht and Lengauer 2004). In the absence of *EDC3,* Dcp1- and Dcp2-defective strains show lower decapping levels (Kshirsagar and Parker 2004). These effects of Edc3 can be explained by its direct interaction with Dcp2 (Harigaya et al. 2010; Nissan et al. 2010).

While Edc3 participates in general decapping, it has an essential role in specific degradation mechanism that ensure autoregulation of *RPS28B* and *YRA1* expression. The presence of a conserved stem-loop structure in the long 3´ UTR of the transcript for the ribosomal protein Rps28b triggers rapid mRNA decay through a mechanism that bypasses deadenylation and directly activates decapping (Badis et al. 2004). *YRA1* auto-regulation occurs by a mechanism in which the protein Yra1 acts to inhibit its own pre-mRNA splicing. The degradation of the unspliced

precursor is stimulated in the presence of Edc3 (Dong et al. 2007). Initial deadenylation is thus not an absolute prerequisite for decapping activation, at least in a few known cases.

The Lsm Complex: RNA Chaperone and Decapping Activator

The Lsm proteins are characterized by the presence of two Sm motifs that are similar to protein sequences found in snRNP particles components. A search for like-Sm domains identified a family of related Lsm proteins (Fromont-Racine et al. 1997; Salgado-Garrido et al. 1999). Two-hybrid screens with each Lsm protein revealed that they were highly connected to each other. These screens also indicated potential roles for the Lsm proteins in two different processes, the nuclear splicing pathway and cytoplasmic mRNA degradation involving Dcp1, Dcp2, Pat1, Xrn1, and Yel015 later known as Edc3/Lsm16 (Fromont-Racine et al. 2000). Lsm1 was shown to facilitate mRNA decapping (Boeck et al. 1998). Affinity purifications and functional experiments demonstrated that the Lsm1–7 cytoplasmic complex was involved in mRNA degradation whereas the Lsm2–8 nuclear complex was involved in splicing. In *lsm* mutant strains, lsm1 to lsm7, mRNAs are still capped but their 3´ ends are oligoadenylated (10–12 residues) indicating that the Lsm complex was required for mRNA decapping (Bouveret et al. 2000; Tharun and Parker 2001; Tharun et al. 2000).

The Lsm1–7 and Lsm2–8 complexes form ring-shaped heptameric assemblies, which directly bind to the 3´ end of mRNAs. In vitro experiments revealed that Lsm complexes preferentially bind oligoadenylated rather than polyadenylated mRNA and that the presence of a U-rich stretch of nucleotides near the 3´ end facilitates the binding (Chowdhury et al. 2007). Binding of Lsm1–7 to the RNA protects the 3´ end of an mRNA from the degradation by the exosome (Chowdhury et al. 2007; He and Parker 2001). Therefore, Lsm complexes could have a role as both enhancers of decapping and protectors of oligoadenylated 3´ end trimming. Unlike many of the factors described here that are specific for eukaryotes, Lsms have equivalents in prokaryotes: the Hfq proteins. Hfq form multimeric rings that bind and affect the function and stability of many noncoding regulatory RNAs in bacteria (for review, see De Lay et al. 2013).

Pat1, Dhh1, and Scd6: Linking Translation and mRNA Decay

Translation and mRNA decay are tightly linked through the action of proteins that have roles in both processes. Three such proteins stand out: Pat1, Dhh1, a DEAD box helicase, and Scd6 that is endowed with an Sm domain variant and was also called Lsm13 (Albrecht and Lengauer 2004). These proteins affect both mRNA degradation and translational repression (Coller and Parker 2005). Furthermore, these factors as well as most of the factors involved in mRNA decapping co-localize with translation repression proteins in P bodies (for review, see Eulalio et al. 2007a).

Pat1 is a key player in mRNA degradation because it serves as a scaffold for decapping and activator of decapping components (Nissan et al. 2010). The N-terminal part of the protein interacts with Dhh1 whereas the middle and C-terminal domains interact with the Lsm complex, Dcp1/2, Xrn1, and the Ccr4/Not complex. Pat1 was originally identified as interacting with all the Lsm proteins in two-hybrid screens (Fromont-Racine et al. 2000). Surprisingly, the binding of Pat1 (via the C-terminal part of Pat1) to the cytoplasmic Lsm complex is not provided by Lsm1, which is the specific cytoplasmic subunit, but by Lsm2 and Lsm3, as shown by recent structural work (Sharif and Conti 2013; Wu et al. 2013). Mutations in the C-terminal domain of Pat1 or in the C-terminal helix of Lsm2 or in the N-terminal helix of Lsm3, which impede the interactions between Pat1 and the Lsm complex affect mRNA decapping and 3´ to 5´ degradation (Wu et al. 2013).

The DEAD box helicase Dhh1 functions in mRNA decapping and interacts with both decapping and deadenylase complexes (Coller et al. 2001). Dhh1 binds to RNA and also associates with Edc3 or the N-terminal part of Pat1. There is a competition between Pat1 and Edc3 for binding to Dhh1, which interferes with the RNA-binding capacity of Dhh1. Competition for binding could thus lead to variable remodeling of the corresponding mRNP (Sharif et al. 2013). It was proposed that Dhh1 is especially present on slowly translated mRNA molecules and that slow translation could favor mRNA decapping (Sweet et al. 2012).

In addition to its Sm domain, Scd6 contains an RGG motif that allows its association with the eIF4G subunit of eIF4F translation initiation complex and participates to translation repression. Edc3 and Scd6 compete through their Lsm motifs for the same HLM domains of Dcp2 (Fromm et al. 2012). In vitro decapping assays revealed that both Edc3 and Scd6 are able to stimulate Dcp1/Dcp2 decapping activity. However, Scd6 has a relatively low affinity for Dcp2- and Scd6-mediated activation is weak. Moreover, since Dcp2 contains several HLM motifs, it can bind concomitantly to different partners and it is not clear in which physiologically relevant conditions the competition between different binders that can affect decapping is important.

Additional Factors in mRNA Decay: Pbp1, Pbp4, and Lsm12

Pbp1, Pab1-binding protein 1, was identified through a two-hybrid screen in association with Pab1 (Mangus et al. 1998). The same group identified later Pbp4, a factor that interacts with Pbp1 (Mangus et al. 2004a). Together with Lsm12, identified through its Sm motif, Pbp1 and Pbp4 are part of the same complex that interacts with ribosomes (Fleischer et al. 2006). The absence of any of these three factors has a negative effect on growth of yeast strains that also contain a deletion of *EDC3* (Decourty et al. 2008). Moreover, these factors colocalize together with Pab1 in stress granules in glucose-deprived cells (Shah et al. 2013). Pbp1 promotes the formation of stress granules that contain Lsm12 and Pbp4, whereas Pbp4 has no effect on the stress granules containing Lsm12 or Pbp1 (Swisher and Parker

2010). Pbp1 was also proposed to be involved in the regulation of poly(A) length through its interaction with Pab1 that could inhibit the recruitment of the Pan2/3 complex (Mangus et al. 2004b). While the Pbp1/Pbp4/Lsm12 complex is functionally linked with mRNA decay, the cellular role of these proteins remains unclear.

The 5′ to 3′ Degradation of Unprotected RNA: Xrn1

Xrn1 is the major 5′ to 3′ exonuclease in the cytoplasm and hydrolyses RNA that starts with a 5′ monophosphate. Pioneer work leading to the discovery and initial characterization of Xrn1 came mostly from A. Stevens laboratory. The existence of a 5′ to 3′ enzymatic activity in eukaryotic cells was first suspected at the end of the 1970s because it was observed that uncapped mRNA were hydrolyzed by crude extracts or by purified enzymatic yeast fractions whereas capped RNA were not (Furuichi et al. 1977; Stevens 1978). Ten years later, the gene coding for Xrn1 was cloned from yeast (Larimer and Stevens, 1990). While yeast cells could adapt to the absence of the gene, their growth rate was severely affected. It was later shown that the absence of Xrn1 leads to accumulation of poly(A) deficient mRNA that lack the cap structure (Hsu and Stevens 1993). Moreover, looking at endogenous mRNA in pulse-chase experiments highlighted a decrease of the mRNA turnover rate in *xrn1* mutants, which was the first evidence for a global role of Xrn1 in mRNA decay. The use of mRNA reporters containing a stretch of Gs (guanosine residues), that Xrn1 cannot degrade, allowed a description of RNA degradation species, leading to current models for mRNA decay (Muhlrad et al. 1994).

RNA sequencing performed with an *xrn1*-deficient strain identified a subgroup of noncoding transcripts, which were called XUT for Xrn1-sensitive unstable transcripts (van Dijk et al. 2011), probably derived from pervasive transcription products that escape to the cytoplasm. In addition to a role of Xrn1 in mRNA turnover, the enzyme is thus also very important for its clearing function for cytoplasmic noncoding RNA.

Xrn1 is involved in lithium toxicity in yeast. It has been shown that lithium inhibits Hal2, an enzyme that converts adenosine 3′,5′ bisphosphate (pAp) into AMP. The accumulation of pAp inhibits Xrn1 and the RNase MRP (Dichtl et al. 1997). Whether these effects in yeast could have a counterpart in the mechanism of action of lithium salts, an effective treatment of psychiatric disorders in humans, remains unknown. However, lithium treatment of yeast cells is an effective tool for the study of otherwise unstable RNA species that are substrates of the ribonucleases.

Recycling of RNA Degradation Byproducts: Dcs1/Dcs2

When a capped RNA is degraded from 3′ to 5′, the last product of the enzymatic hydrolysis is the dinucleotide m7GpppN or oligonucleotides of the form m7GpppNNN. A "scavenger" mRNA decapping enzyme that can degrade these end products of mRNA decay, DcpS, was first identified in mammalian cells (Liu et al. 2002; Wang and Kiledjian 2001). The enzyme contains a HIT motif (His-X-His-X-His-X, where X is a hydrophobic amino acid) essential for the cap hydrolysis activity. Recombinant DcpS is able to hydrolyze analogs of methylated cap or capped mRNA inferior in length to 10 nucleotides, suggesting that DcpS is unable to bind intact mRNA but degrade the residual cap structure after 3′ to 5′ exosome degradation. A sequence search in *S. cerevisiae* database identified two orthologs of human DcpS, Dcs1/Dcs2 (Liu et al. 2002). Despite the strong homology between Dcs1 and Dcs2, only Dcs1 seem to have a catalytic activity. Both proteins form a heterodimer and it was suggested that Dcs2 could be a modulator of Dcs1 activity (Malys et al. 2004). Deletion of Dcs1 leads to an accumulation of uncapped mRNA, which is the result of a decrease of 5′ to 3′ exonuclease activity showing that Dcs1/Dcs2 are not only involved in the clearance of the cap structure but also participate in the 5′ to 3′ mRNA decay by facilitating Xrn1 activity (Liu and Kiledjian 2005). It has been recently shown by in vitro assays that Dcs1 directly activates Xrn1 and that Dcs1 is an in vivo cofactor of Xrn1 important for respiration in yeast (Sinturel et al. 2012).

P Bodies are Large Aggregates of Proteins Involved in RNA Decay

An estimated 15 % of yeast proteins form aggregates in the cytoplasm of starved cells, as discovered from systematic analyses of GFP fusion yeast strain collections (Narayanaswamy et al. 2009; Noree et al. 2010; as reviewed by O'Connell et al. 2012). Many of the described aggregation-prone proteins are involved in glucose, amino acid, or nucleotide metabolism. Proteins involved in mRNA decay, like the decapping enzyme Dcp2 and exonuclease Xrn1 had been also known for some time to form visible foci in the cytoplasm of mammalian (Bashkirov et al. 1997; van Dijk et al. 2002; Ingelfinger et al. 2002) and yeast cells (Sheth and Parker 2003). Several experiments were used to propose that the Dcp/Xrn1 aggregates, called processing bodies or P bodies were sites where mRNA decapping and 5′ to 3′ degradation occurs (see also Cougot et al. 2004 for mammalian cells experiments). Treatment of cells with cycloheximide, which blocks translation, is followed by a redistribution of P body components, like Dcp1, to the entire cytoplasm (Sheth and Parker 2003).

While P bodies could be places of mRNA degradation, the bulk of mRNA degradation was proposed to occur in polysome complexes (Hu et al. 2009). In addition, P body formation was found to be uncoupled from mRNA degradation

both in yeast and in *D. melanogaster* cells (Decker et al. 2007; Eulalio et al. 2007b; Sweet et al. 2007). The number of proteins capable of entering P bodies has increased recently as a result of a systematic search for RNA-binding proteins (Mitchell et al. 2013). The list of proteins present in distinct cytoplasmic foci when cells are deprived of glucose is long and includes: Dhh1, Eap1, Gis2, Hek2, Pat1, Pbp1, Pub1, Puf2-3-4-5, Sro9, Tif4631, Upf1, Upf3, and Xrn1 among others. A historical perspective of P body discovery thoroughly describes both the complex protein composition of these aggregates and the relationship with other cytoplasmic large aggregates like stress granules (Jain and Parker 2013).

The main difference between stress granules and P bodies is that they form as a results of a different stress and contain 40S ribosomal subunit components together with translation initiation factors. Due to this difference in composition, it was proposed that stress granules serve as sites of mRNA storage (reviewed in Yamasaki and Anderson 2008).

P bodies could be a result of the propensity of RNA-binding proteins to aggregate and their formation is increased by stress. It has been proposed recently that Pat1, one of the factors that could link mRNA translation and decay, could be the promoter of P body formation under glucose starvation (Ramachandran et al. 2011; Shah et al. 2013). Protein kinase A (PKA) signaling in yeast is important for the response of cells to glucose concentration. PKA is activated in response to glucose and leads to major changes in the yeast transcriptome (Zaman et al. 2009). Pat1 phosphorylation by PKA was shown to affect P body formation and point mutations of the phosphorylated residues influence the number of foci and their dynamics (Shah et al. 2013). At least another signaling pathway, involving protein kinases Pkh1/2 and Pkc1 affect P body formation and mRNA decay (Luo et al. 2011). The role of P bodies, how they form, and the extent to which they serve as foci of RNA degradation or storage remains under very active investigation.

Mitochondrial mRNA Degradation

A relatively small group of 19 protein-coding RNAs in *S. cerevisiae* are synthesized in mitochondria (Turk et al. 2013) and are subject to intraorganellar decay through mechanisms that are radically different from the nuclear derived mRNAs. Formation of mature mRNA sequences depends on processing of multigene transcripts. The mRNAs are not polyadenylated but can contain large untranslated regions. An interesting feature of mitochondrial mRNA 3′ end is the presence of an A-rich sequence called dodecamer 5′-AAUAAUAUUCUU-3′ that serves both as an endonucleolytic mark and as a protective sequence (Hofmann et al. 1993; Osinga et al. 1984).

mRNA degradation is mainly achieved by a complex named mitochondrial degradosome (mtEXO) that has two components: Dss1, a 3′ to 5′ exonuclease related to the RNAse II-like family (Dmochowska et al. 1995; Min et al. 1993) and Suv3, an ATP-dependent RNA helicase related to the Ski2 DExH/D superfamily

(Stepien et al. 1992). Both components of the mtEXO are coded by the nuclear genome. The absence of Suv3 leads to a respiratory phenotype and to the accumulation of aberrant mitochondrial RNA. Point mutations in the gene for the mitochondrial RNA polymerase (Rpo41) or its essential cofactor Mtf1 partially restore the phenotype of a *suv3*Δ strain. It was proposed that maintenance of the balance between mRNA synthesis and degradation is essential for mitochondrial function (Rogowska et al. 2006). In contrast to the related bacterial degradosomes (reviewed in Bandyra et al. 2013), mtEXO degrades mRNAs that are not polyadenylated in *S. cerevisiae*.

Mitochondrial gene regulation seems to be highly different among eukaryotes. As an example, in *Arabidopsis thaliana*, mitochondrial mRNAs are polyadenylated and it was recently shown that two proteins regulate the poly(A) tail length. Interestingly, expression of the poly(A) polymerase *AGS1* from *A. thaliana* in yeast results in polyadenylation of the *COX3* mitochondrial mRNA, which is not normally polyadenylated (Hirayama et al. 2013). In *S. pombe*, the equivalent of mtEXO is composed of two proteins, Pah1 for the helicase and Par1 for the RNase, but RNA degradation is significantly different from *S. cerevisiae*. *SUV3* overexpression is able to restore the defect of *pah1*Δ whereas Dss1 is not functional in a *par1*Δ strain. In *S. pombe*, this complex was proposed to be mainly involved in 3′ end mRNA maturation and was accordingly named a "processosome" (Hoffmann et al. 2008).

The existence of a 5′ to 3′ degradation pathway for mitochondrial mRNAs is still questionable. Since mitochondrial mRNAs are synthesized as multigenic transcripts, their 5′ end needs processing to become mature. In the absence of Pet127, precursor mRNAs accumulate. Moreover, when the 5′ ends of intermediate mRNAs are not protected, they accumulate in *pet127*Δ mutant strain (Wiesenberger and Fox 1997). These results suggest that processing and degradation could be coupled. Pet127 is a potential candidate for a 5′ to 3′ mRNA degradation factor, but there is no direct evidence of an enzymatic function of the protein, which does not possess any detectable RNase signature. However, overexpression of *PET127* can suppress the deletion of *SUV3* or *DSS1* suggesting that 5′ to 3′ and the 3′ to 5′ redundant pathways could exist in mitochondria (Wegierski et al. 1998).

Conclusion

While many factors involved in mRNA degradation and turnover are now known to exist, how they collaborate, on which substrates they act and under what environmental conditions, remain open questions. It will be interesting to see to what extent general models of mRNA decay proposed on the basis of studies of individual reporters will stand the test of time and the "storm" of large-scale data that are growing exponentially.

Acknowledgments We thank Alain Jacquier and Frank Feuerbach for criticism of the manuscript. Funding was provided by the Institut Pasteur, the CNRS, and the French Agence Nationale de la Recherche (ANR-2011-BSV6-011-02 and ANR-2008-JCJC-0019-01/GENO-GIM).

References

Albrecht M, Lengauer T (2004) Novel Sm-like proteins with long C-terminal tails and associated methyltransferases. FEBS Lett 569:18–26

Amberg DC, Goldstein AL, Cole CN (1992) Isolation and characterization of RAT1: an essential gene of *Saccharomyces cerevisiae* required for the efficient nucleocytoplasmic trafficking of mRNA. Genes Dev 6:1173–1189

Amrani N, Ghosh S, Mangus DA, Jacobson A (2008) Translation factors promote the formation of two states of the closed-loop mRNP. Nature 453:1276–1280

Anderson JT, Wilson SM, Datar KV, Swanson MS (1993) NAB2: a yeast nuclear polyadenylated RNA-binding protein essential for cell viability. Mol Cell Biol 13:2730–2741

Araki Y, Takahashi S, Kobayashi T, Kajiho H, Hoshino S, Katada T (2001) Ski7p G protein interacts with the exosome and the Ski complex for 3´-to-5´ mRNA decay in yeast. EMBO J 20:4684–4693

Arava Y, Wang Y, Storey JD, Liu CL, Brown PO, Herschlag D (2003) Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. Proc Natl Acad Sci USA 100:3889–3894

Arigo JT, Carroll KL, Ames JM, Corden JL (2006) Regulation of yeast NRD1 expression by premature transcription termination. Mol Cell 21:641–651

Arraiano CM, Mauxion F, Viegas SC, Matos RG, Séraphin B (2013) Intracellular ribonucleases involved in transcript processing and decay: precision tools for RNA. Biochim Biophys Acta 1829:491–513

Badis G, Saveanu C, Fromont-Racine M, Jacquier A (2004) Targeted mRNA degradation by deadenylation-independent decapping. Mol Cell 15:5–15

Baer BW, Kornberg RD (1980) Repeating structure of cytoplasmic poly(A)-ribonucleoprotein. Proc Natl Acad Sci USA 77:1890–1892

Baer BW, Kornberg RD (1983) The protein responsible for the repeating structure of cytoplasmic poly(A)-ribonucleoprotein. J Cell Biol 96:717–721

Bandyra KJ, Bouvier M, Carpousis AJ, Luisi BF (2013) The social fabric of the RNA degradosome. Biochim Biophys Acta 1829:514–522

Bashkirov VI, Scherthan H, Solinger JA, Buerstedde JM, Heyer WD (1997) A mouse cytoplasmic exoribonuclease (mXRN1p) with preference for G4 tetraplex substrates. J Cell Biol 136:761–773

Basquin J, Roudko VV, Rode M, Basquin C, Séraphin B, Conti E (2012) Architecture of the nuclease module of the yeast Ccr4-not complex: the Not1-Caf1-Ccr4 interaction. Mol Cell 48:207–218

Beelman CA, Stevens A, Caponigro G, LaGrandeur TE, Hatfield L, Fortner DM, Parker R (1996) An essential component of the decapping enzyme required for normal rates of mRNA turnover. Nature 382:642–646

Boeck R, Tarun S Jr, Rieger M, Deardorff JA, Müller-Auer S, Sachs AB (1996) The yeast Pan2 protein is required for poly(A)-binding protein-stimulated poly(A)-nuclease activity. J Biol Chem 271:432–438

Boeck R, Lapeyre B, Brown CE, Sachs AB (1998) Capped mRNA degradation intermediates accumulate in the yeast spb8-2 mutant. Mol Cell Biol 18:5062–5072

Borja MS, Piotukh K, Freund C, Gross JD (2011) Dcp1 links coactivators of mRNA decapping to Dcp2 by proline recognition. RNA 17:278–290

Bousquet-Antonelli C, Presutti C, Tollervey D (2000) Identification of a regulated pathway for nuclear pre-mRNA turnover. Cell 102:765–775

Bouveret E, Rigaut G, Shevchenko A, Wilm M, Séraphin B (2000) A Sm-like protein complex that participates in mRNA degradation. EMBO J 19:1661–1671

Bregman A, Avraham-Kelbert M, Barkai O, Duek L, Guterman A, Choder M (2011) Promoter elements regulate cytoplasmic mRNA Decay. Cell 147:1473–1483

Brown CE, Sachs AB (1998) Poly(A) tail length control in *Saccharomyces cerevisiae* occurs by message-specific deadenylation. Mol Cell Biol 18:6548–6559

Brown CE, Tarun SZ Jr, Boeck R, Sachs AB (1996) PAN3 encodes a subunit of the Pab1p-dependent poly(A) nuclease in *Saccharomyces cerevisiae*. Mol Cell Biol 16:5744–5753

Buchan JR, Muhlrad D, Parker R (2008) P bodies promote stress granule assembly in Saccharomyces cerevisiae. J Cell Biol 183:441–455

Burger K, Mühl B, Kellner M, Rohrmoser M, Gruber-Eber A, Windhager L, Friedel CC, Dölken L, Eick D (2013) 4-thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. RNA Biol 10

Butler JS, Mitchell P (2011) Rrp6, rrp47 and cofactors of the nuclear exosome. Adv Exp Med Biol 702:91–104

Callahan KP, Butler JS (2010) TRAMP complex enhances RNA degradation by the nuclear exosome component Rrp6. J Biol Chem 285:3540–3547

Chang JH, Jiao X, Chiba K, Oh C, Martin CE, Kiledjian M, Tong L (2012) Dxo1 is a new type of eukaryotic enzyme with both decapping and 5′-3′ exoribonuclease activity. Nat Struct Mol Biol 19:1011–1017

Chowdhury A, Mukhopadhyay J, Tharun S (2007) The decapping activator Lsm1p-7p-Pat1p complex has the intrinsic ability to distinguish between oligoadenylated and polyadenylated RNAs. RNA 13:998–1016

Cleary MD, Meiering CD, Jan E, Guymon R, Boothroyd JC (2005) Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. Nat Biotechnol 23:232–237

Collart MA, Panasenko OO (2012) The Ccr4–not complex. Gene 492:42–53

Coller J, Parker R (2005) General translational repression by activators of mRNA decapping. Cell 122:875–886

Coller JM, Tucker M, Sheth U, Valencia-Sanchez MA, Parker R (2001) The DEAD box helicase, Dhh1p, functions in mRNA decapping and interacts with both the decapping and deadenylase complexes. RNA 7:1717–1727

Costello JL, Stead JA, Feigenbutz M, Jones RM, Mitchell P (2011) The C-terminal region of the exosome-associated protein Rrp47 is specifically required for box C/D small nucleolar RNA 3′-maturation. J Biol Chem 286:4535–4543

Cougot N, Babajko S, Séraphin B (2004) Cytoplasmic foci are sites of mRNA decay in human cells. J Cell Biol 165:31–40

Daugeron MC, Mauxion F, Séraphin B (2001) The yeast POP2 gene encodes a nuclease involved in mRNA deadenylation. Nucleic Acids Res 29:2448–2455

Decker CJ, Teixeira D, Parker R (2007) Edc3p and a glutamine/asparagine-rich domain of Lsm4p function in processing body assembly in Saccharomyces cerevisiae. J Cell Biol 179:437–449

Decourty L, Saveanu C, Zemam K, Hantraye F, Frachon E, Rousselle J-C, Fromont-Racine M, Jacquier A (2008) Linking functionally related genes by sensitive and quantitative characterization of genetic interaction profiles. Proc Natl Acad Sci USA 105:5821–5826

Deshmukh MV, Jones BN, Quang-Dang D-U, Flinders J, Floor SN, Kim C, Jemielity J, Kalek M, Darzynkiewicz E, Gross JD (2008) mRNA decapping is promoted by an RNA-binding channel in Dcp2. Mol Cell 29:324–336

Dichtl B, Stevens A, Tollervey D (1997) Lithium toxicity in yeast is due to the inhibition of RNA processing enzymes. EMBO J 16:7184–7195

Van Dijk E, Cougot N, Meyer S, Babajko S, Wahle E, Séraphin B (2002) Human Dcp2: a catalytically active mRNA decapping enzyme located in specific cytoplasmic structures. EMBO J 21:6915–6924

Van Dijk EL, Chen CL, d´ Aubenton-Carafa Y, Gourvennec S, Kwapisz M, Roche V, Bertrand C, Silvain M, Legoix-Né P, Loeillet S et al (2011) XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. Nature 475:114–117

Dmochowska A, Golik P, Stepien PP (1995) The novel nuclear gene DSS-1 of *Saccharomyces cerevisiae* is necessary for mitochondrial biogenesis. Curr Genet 28:108–112

Dölken L, Ruzsics Z, Rädle B, Friedel CC, Zimmer R, Mages J, Hoffmann R, Dickinson P, Forster T, Ghazal P et al (2008) High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. RNA 14:1959–1972

Dong S, Li C, Zenklusen D, Singer RH, Jacobson A, He F (2007) YRA1 autoregulation requires nuclear export and cytoplasmic Edc3p-mediated degradation of its pre-mRNA. Mol Cell 25:559–573

Dori-Bachash M, Shalem O, Manor YS, Pilpel Y, Tirosh I (2012) Widespread promoter-mediated coordination of transcription and mRNA degradation. Genome Biol 13:R114

Drinnenberg IA, Fink GR, Bartel DP (2011) Compatibility with killer explains the rise of RNAi-deficient fungi. Science 333:1592

Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, De Montigny J, Marck C, Neuvéglise C, Talla E et al (2004) Genome evolution in yeasts. Nature 430:35–44

Dunckley T, Parker R (1999) The DCP2 protein is required for mRNA decapping in *Saccharomyces cerevisiae* and contains a functional MutT motif. EMBO J 18:5411–5422

Dunckley T, Tucker M, Parker R (2001) Two related proteins, Edc1p and Edc2p, stimulate mRNA decapping in Saccharomyces cerevisiae. Genetics 157:27–37

Dziembowski A, Lorentzen E, Conti E, Séraphin B (2007) A single subunit, Dis3, is essentially responsible for yeast exosome core activity. Nat Struct Mol Biol 14:15–22

Engel SR, Cherry JM (2013) The new modern era of yeast genomics: community sequencing and the resulting annotation of multiple *Saccharomyces cerevisiae* strains at the Saccharomyces Genome Database. Database J Biol, Databases Curation 2013

Ensinger MJ, Martin SA, Paoletti E, Moss B (1975) Modification of the 5´-terminus of mRNA by soluble guanylyl and methyl transferases from vaccinia virus. Proc Natl Acad Sci USA 72:2525–2529

Eulalio A, Behm-Ansmant I, Izaurralde E (2007a) P bodies: at the crossroads of post-transcriptional pathways. Nat Rev Mol Cell Biol 8:9–22

Eulalio A, Behm-Ansmant I, Schweizer D, Izaurralde E (2007b) P-body formation is a consequence, not the cause, of RNA-mediated gene silencing. Mol Cell Biol 27:3970–3981

Fleischer TC, Weaver CM, McAfee KJ, Jennings JL, Link AJ (2006) Systematic identification and functional screens of uncharacterized proteins associated with eukaryotic ribosomal complexes. Genes Dev 20:1294–1307

Foat BC, Houshmandi SS, Olivas WM, Bussemaker HJ (2005) Profiling condition-specific, genome-wide regulation of mRNA stability in yeast. Proc Natl Acad Sci USA 102:17675–17680

Fromm SA, Truffault V, Kamenz J, Braun JE, Hoffmann NA, Izaurralde E, Sprangers R (2012) The structural basis of Edc3- and Scd6-mediated activation of the Dcp1:Dcp2 mRNA decapping complex. EMBO J 31:279–290

Fromont-Racine M, Rain JC, Legrain P (1997) Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. Nat Genet 16:277–282

Fromont-Racine M, Mayes AE, Brunet-Simon A, Rain JC, Colley A, Dix I, Decourty L, Joly N, Ricard F, Beggs JD et al (2000) Genome-wide protein interaction screens reveal functional networks involving Sm-like proteins. Yeast 17:95–110

Furuichi Y, LaFiandra A, Shatkin AJ (1977) 5´-Terminal structure and mRNA stability. Nature 266:235–239

Garland W, Feigenbutz M, Turner M, Mitchell P (2013) Rrp47 functions in RNA surveillance and stable RNA processing when divorced from the exoribonuclease and exosome-binding domains of Rrp6. RNA 19(12):1659–1668

Gerber AP, Herschlag D, Brown PO (2004) Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. PLoS Biol 2: e79

Goebels C, Thonn A, Gonzalez-Hilarion S, Rolland O, Moyrand F, Beilharz TH, Janbon G (2013) Introns Regulate Gene Expression in Cryptococcus neoformans in a Pab2p Dependent Pathway. PLoS Genet 9:e1003686

Goldstrohm AC, Hook BA, Seay DJ, Wickens M (2006) PUF proteins bind Pop2p to regulate messenger RNAs. Nat Struct Mol Biol 13:533–539

Goldstrohm AC, Seay DJ, Hook BA, Wickens M (2007) PUF protein-mediated deadenylation is catalyzed by Ccr4p. J Biol Chem 282:109–114

Granneman S, Kudla G, Petfalski E, Tollervey D (2009) Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. Proc Natl Acad Sci USA 106:9613–9618

Green DM, Marfatia KA, Crafton EB, Zhang X, Cheng X, Corbett AH (2002) Nab2p is required for poly(A) RNA export in *Saccharomyces cerevisiae* and is regulated by arginine methylation via Hmt1p. J Biol Chem 277:7752–7760

Grigull J, Mnaimneh S, Pootoolal J, Robinson MD, Hughes TR (2004) Genome-wide analysis of mRNA stability using transcription inhibitors and microarrays reveals posttranscriptional control of ribosome biogenesis factors. Mol Cell Biol 24:5534–5547

Gu M, Rajashankar KR, Lima CD (2010) Structure of the *Saccharomyces cerevisiae* Cet1-Ceg1 mRNA capping apparatus. Struct Lond Engl 1993(18):216–227

Gudipati RK, Neil H, Feuerbach F, Malabat C, Jacquier A (2012) The yeast RPL9B gene is regulated by modulation between two modes of transcription termination. EMBO J 31:2427–2437

Halbach F, Reichelt P, Rode M, Conti E (2013) The yeast ski complex: crystal structure and RNA channeling to the exosome complex. Cell 154:814–826

Halbeisen RE, Scherrer T, Gerber AP (2009) Affinity purification of ribosomes to access the translatome. Methods San Diego Calif 48:306–310

Harigaya Y, Jones BN, Muhlrad D, Gross JD, Parker R (2010) Identification and analysis of the interaction between Edc3 and Dcp2 in Saccharomyces cerevisiae. Mol Cell Biol 30:1446–1456

He W, Parker R (2001) The yeast cytoplasmic LsmI/Pat1p complex protects mRNA 3′ termini from partial degradation. Genetics 158:1445–1455

Hector RE, Nykamp KR, Dheur S, Anderson JT, Non PJ, Urbinati CR, Wilson SM, Minvielle-Sebastia L, Swanson MS (2002) Dual requirement for yeast hnRNP Nab2p in mRNA poly(A) tail length control and nuclear export. EMBO J 21:1800–1810

Hirayama T, Matsuura T, Ushiyama S, Narusaka M, Kurihara Y, Yasuda M, Ohtani M, Seki M, Demura T, Nakashita H et al (2013) A poly(A)-specific ribonuclease directly regulates the poly(A) status of mitochondrial mRNA in Arabidopsis. Nat Commun 4:2247

Hoffmann B, Nickel J, Speer F, Schafer B (2008) The 3′ ends of mature transcripts are generated by a processosome complex in fission yeast mitochondria. J Mol Biol 377:1024–1037

Hofmann TJ, Min J, Zassenhaus HP (1993) Formation of the 3′ end of yeast mitochondrial mRNAs occurs by site-specific cleavage two bases downstream of a conserved dodecamer sequence. Yeast 9:1319–1330

Hogan DJ, Riordan DP, Gerber AP, Herschlag D, Brown PO (2008) Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. PLoS Biol 6:e255

Holub P, Lalakova J, Cerna H, Pasulka J, Sarazova M, Hrazdilova K, Arce MS, Hobor F, Stefl R, Vanacova S (2012) Air2p is critical for the assembly and RNA-binding of the TRAMP complex and the KOW domain of Mtr4p is crucial for exosome activation. Nucleic Acids Res 40:5679–5693

Houseley J, LaCava J, Tollervey D (2006) RNA-quality control by the exosome. Nat Rev Mol Cell Biol 7:529–539

Hsu CL, Stevens A (1993) Yeast cells lacking 5′–>3′ exoribonuclease 1 contain mRNA species that are poly(A) deficient and partially lack the 5′ cap structure. Mol Cell Biol 13:4826–4835

Hu W, Sweet TJ, Chamnongpol S, Baker KE, Coller J (2009) Co-translational mRNA decay in Saccharomyces cerevisiae. Nature 461:225–229

Iglesias N, Tutucci E, Gwizdek C, Vinciguerra P, Von Dach E, Corbett AH, Dargemont C, Stutz F (2010) Ubiquitin-mediated mRNP dynamics and surveillance prior to budding yeast mRNA export. Genes Dev 24:1927–1938

Ingelfinger D, Arndt-Jovin DJ, Lührmann R, Achsel T (2002) The human LSm1-7 proteins colocalize with the mRNA-degrading enzymes Dcp1/2 and Xrnl in distinct cytoplasmic foci. RNA 8:1489–1501

Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324:218–223

Jackson JS Jr, Houshmandi SS, Lopez Leban F, Olivas WM (2004) Recruitment of the Puf3 protein to its mRNA target for regulation of mRNA decay in yeast. RNA 10:1625–1636

Jain S, Parker R (2013) The discovery and analysis of P Bodies. Adv Exp Med Biol 768:23–43

Jensen TH, Jacquier A, Libri D (2013) Dealing with pervasive transcription. Mol Cell 52:473–484

Jiao X, Xiang S, Oh C, Martin CE, Tong L, Kiledjian M (2010) Identification of a quality-control mechanism for mRNA 5′-end capping. Nature 467:608–611

Jimenez A, Tipper DJ, Davies J (1973) Mode of Action of Thiolutin, an Inhibitor of Macromolecular Synthesis in Saccharomyces cerevisiae. Antimicrob Agents Chemother 3:729–738

Johnson AW (1997) Rat1p and Xrn1p are functionally interchangeable exoribonucleases that are restricted to and required in the nucleus and cytoplasm, respectively. Mol Cell Biol 17:6122–6130

Keene JD (2007) RNA regulons: coordination of post-transcriptional events. Nat Rev Genet 8:533–543

Keene JD, Tenenbaum SA (2002) Eukaryotic mRNPs may represent posttranscriptional operons. Mol Cell 9:1161–1167

Kim Guisbert K, Duncan K, Li H, Guthrie C (2005) Functional specificity of shuttling hnRNPs revealed by genome-wide analysis of their RNA binding profiles. RNA 11:383–393

Klass DM, Scheibe M, Butter F, Hogan GJ, Mann M, Brown PO (2013) Quantitative proteomic analysis reveals concurrent RNA-protein interactions and identifies new RNA-binding proteins in Saccharomyces cerevisiae. Genome Res 23:1028–1038

Kshirsagar M, Parker R (2004) Identification of Edc3p as an enhancer of mRNA decapping in Saccharomyces cerevisiae. Genetics 166:729–739

LaCava J, Houseley J, Saveanu C, Petfalski E, Thompson E, Jacquier A, Tollervey D (2005) RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. Cell 121:713–724

Larimer FW, Stevens A (1990) Disruption of the gene XRN1, coding for a 5′—3′ exoribonuclease, restricts yeast cell growth. Gene 95:85–90

De Lay N, Schu DJ, Gottesman S (2013) Bacterial small RNA-based negative regulation: Hfq and its accomplices. J Biol Chem 288:7996–8003

Lebreton A, Tomecki R, Dziembowski A, Séraphin B (2008) Endonucleolytic RNA cleavage by a eukaryotic exosome. Nature 456:993–996

Legrain P, Rosbash M (1989) Some cis- and trans-acting mutants for splicing target pre-mRNA to the cytoplasm. Cell 57:573–583

Lemieux C, Bachand F (2009) Cotranscriptional recruitment of the nuclear poly(A)-binding protein Pab2 to nascent transcripts and association with translating mRNPs. Nucleic Acids Res 37:3418–3430

Lidschreiber M, Leike K, Cramer P (2013) Cap completion and C-terminal repeat domain kinase recruitment underlie the initiation-elongation transition of RNA polymerase II. Mol Cell Biol 33:3805–3816

Liu H, Kiledjian M (2005) Scavenger Decapping Activity Facilitates 5′ to 3′ mRNA Decay. Mol Cell Biol 25:9764–9772

Liu H, Rodgers ND, Jiao X, Kiledjian M (2002) The scavenger mRNA decapping enzyme DcpS is a member of the HIT family of pyrophosphatases. EMBO J 21:4699–4708

Liu Q, Greimann JC, Lima CD (2006) Reconstitution, activities, and structure of the eukaryotic RNA exosome. Cell 127:1223–1237

Lo TL, Qu Y, Uwamahoro N, Quenault T, Beilharz TH, Traven A (2012) The mRNA decay pathway regulates the expression of the Flo11 adhesin and biofilm formation in Saccharomyces cerevisiae. Genetics 191:1387–1391

Lowell JE, Rudner DZ, Sachs AB (1992) 3´-UTR-dependent deadenylation by the yeast poly(A) nuclease. Genes Dev 6:2088–2099

Luo G, Costanzo M, Boone C, Dickson RC (2011) Nutrients and the Pkh1/2 and Pkc1 protein kinases control mRNA decay and P-body assembly in yeast. J Biol Chem 286:8759–8770

Luo W, Johnson AW, Bentley DL (2006) The role of Rat1 in coupling mRNA 3?-end processing to transcription termination: implications for a unified allosteric-torpedo model. Genes Dev 20:954–965

Makino DL, Baumgärtner M, Conti E (2013a) Crystal structure of an RNA-bound 11-subunit eukaryotic exosome complex. Nature 495:70–75

Makino DL, Halbach F, Conti E (2013b) The RNA exosome and proteasome: common principles of degradation control. Nat Rev Mol Cell Biol 14:654–660

Malecki M, Viegas SC, Carneiro T, Golik P, Dressaire C, Ferreira MG, Arraiano CM (2013) The exoribonuclease Dis3L2 defines a novel eukaryotic RNA degradation pathway. EMBO J 32:1842–1854

Malys N, Carroll K, Miyan J, Tollervey D, McCarthy JEG (2004) The "scavenger" m7G pppX pyrophosphatase activity of Dcs1 modulates nutrient-induced responses in yeast. Nucleic Acids Res 32:3590–3600

Mangus DA, Amrani N, Jacobson A (1998) Pbp1p, a factor interacting with *Saccharomyces cerevisiae* poly(A)-binding protein, regulates polyadenylation. Mol Cell Biol 18:7383–7396

Mangus DA, Smith MM, McSweeney JM, Jacobson A (2004a) Identification of factors regulating poly(A) tail synthesis and maturation. Mol Cell Biol 24:4196–4206

Mangus DA, Evans MC, Agrin NS, Smith M, Gongidi P, Jacobson A (2004b) Positive and negative regulation of poly(A) nuclease. Mol Cell Biol 24:5521–5533

Marshall AN, Montealegre MC, Jiménez-López C, Lorenz MC, van Hoof A (2013) Alternative splicing and subfunctionalization generates functional diversity in fungal proteomes. PLoS Genet 9:e1003376

Matunis MJ, Matunis EL, Dreyfuss G (1993) PUB1: a major yeast poly(A)+RNA-binding protein. Mol Cell Biol 13:6114–6123

Menschaert G, Van Criekinge W, Notelaers T, Koch A, Crappé J, Gevaert K, Van Damme P (2013) Deep proteome coverage based on ribosome profiling aids mass spectrometry-based protein and peptide discovery and provides evidence of alternative translation products and near-cognate translation initiation events. Mol. Cell. Proteomics MCP 12:1780–1790

Miller C, Schwalb B, Maier K, Schulz D, Dümcke S, Zacher B, Mayer A, Sydow J, Marcinowski L, Dölken L et al (2011) Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. Mol Syst Biol 7:458

Milligan L, Decourty L, Saveanu C, Rappsilber J, Ceulemans H, Jacquier A, Tollervey D (2008) A yeast exosome cofactor, Mpp6, functions in RNA surveillance and in the degradation of noncoding RNA transcripts. Mol Cell Biol 28:5446–5457

Min J, Heuertz RM, Zassenhaus HP (1993) Isolation and characterization of an NTP-dependent 3´-exoribonuclease from mitochondria of Saccharomyces cerevisiae. J Biol Chem 268:7350–7357

Mitchell P, Petfalski E, Shevchenko A, Mann M, Tollervey D (1997) The exosome: a conserved eukaryotic RNA processing complex containing multiple 3´–>5´ exoribonucleases. Cell 91:457–466

Mitchell SF, Jain S, She M, Parker R (2013) Global analysis of yeast mRNPs. Nat Struct Mol Biol 20:127–133

Moteki S, Price D (2002) Functional coupling of capping and transcription of mRNA. Mol Cell 10:599–609

Muhlrad D, Decker CJ, Parker R (1994) Deadenylation of the unstable mRNA encoded by the yeast MFA2 gene leads to decapping followed by 5´–>3´ digestion of the transcript. Genes Dev 8:855–866

Munchel SE, Shultzaberger RK, Takizawa N, Weis K (2011) Dynamic profiling of mRNA turnover reveals gene-specific and system-wide regulation of mRNA decay. Mol Biol Cell 22:2787–2795

Narayanaswamy R, Levy M, Tsechansky M, Stovall GM, O'Connell JD, Mirrielees J, Ellington AD, Marcotte EM (2009) Widespread reorganization of metabolic enzymes into reversible assemblies upon nutrient starvation. Proc Natl Acad Sci USA 106:10147–10152

Nissan T, Rajyaguru P, She M, Song H, Parker R (2010) Decapping activators in *Saccharomyces cerevisiae* act by multiple mechanisms. Mol Cell 39:773–783

Nonet M, Scafe C, Sexton J, Young R (1987) Eucaryotic RNA polymerase conditional mutant that rapidly ceases mRNA synthesis. Mol Cell Biol 7:1602–1611

Noree C, Sato BK, Broyer RM, Wilhelm JE (2010) Identification of novel filament-forming proteins in *Saccharomyces cerevisiae* and Drosophila melanogaster. J Cell Biol 190:541–551

O´Connell JD, Zhao A, Ellington AD, Marcotte EM (2012) Dynamic reorganization of metabolic enzymes into intracellular bodies. Annu Rev Cell Dev Biol 28:89–111

Olivas W, Parker R (2000) The Puf3 protein is a transcript-specific regulator of mRNA degradation in yeast. EMBO J 19:6602–6611

Osinga KA, De Vries E, Van der Horst G, Tabak HF (1984) Processing of yeast mitochondrial messenger RNAs at a conserved dodecamer sequence. EMBO J 3:829–834

Parker R (2012) RNA degradation in Saccharomyces cerevisae. Genetics 191:671–702

Pearson EL, Moore CL (2013) Dismantling promoter-driven RNA polymerase II transcription complexes in vitro by the termination factor Rat1. J Biol Chem 288:19750–19759

Pelechano V, Pérez-Ortín JE (2008) The transcriptional inhibitor thiolutin blocks mRNA degradation in yeast. Yeast 25:85–92

Pelechano V, Chávez S, Pérez-Ortín JE (2010) A complete set of nascent transcription rates for yeast genes. PLoS ONE 5:e15442

Perez-Ortin JE, de Miguel-Jimenez L, Chavez S (2012) Genome-wide studies of mRNA synthesis and degradation in eukaryotes. Biochim Biophys Acta-Gene Regul Mech 1819:604–615

Porrua O, Hobor F, Boulay J, Kubicek K, D´Aubenton-Carafa Y, Gudipati RK, Stefl R, Libri D (2012) In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination. EMBO J 31:3935–3948

Qiu C, Kershner A, Wang Y, Holley CP, Wilinski D, Keles S, Kimble J, Wickens M, Hall TMT (2012) Divergence of Pumilio/fem-3 mRNA binding factor (PUF) protein specificity through variations in an RNA-binding pocket. J Biol Chem 287:6949–6957

Quenault T, Lithgow T, Traven A (2011) PUF proteins: repression, activation and mRNA localization. Trends Cell Biol 21:104–112

Ramachandran V, Shah KH, Herman PK (2011) The cAMP-dependent protein kinase signaling pathway is a key regulator of P body foci formation. Mol Cell 43:973–981

Read RL, Martinho RG, Wang S-W, Carr AM, Norbury CJ (2002) Cytoplasmic poly(A) polymerases mediate cellular responses to S phase arrest. Proc Natl Acad Sci USA 99:12079–12084

Rissland OS, Norbury CJ (2009) Decapping is preceded by 3´ uridylation in a novel pathway of bulk mRNA turnover. Nat Struct Mol Biol 16:616–623

Rogowska AT, Puchta O, Czarnecka AM, Kaniak A, Stepien PP, Golik P (2006) Balance between transcription and RNA degradation is vital for *Saccharomyces cerevisiae* mitochondria: reduced transcription rescues the phenotype of deficient RNA degradation. Mol Biol Cell 17:1184–1193

Roth KM, Wolf MK, Rossi M, Butler JS (2005) The nuclear exosome contributes to autogenous control of NAB2 mRNA levels. Mol Cell Biol 25:1577–1585

Roth KM, Byam J, Fang F, Butler JS (2009) Regulation of NAB2 mRNA 3´-end formation requires the core exosome and the Trf4p component of the TRAMP complex. RNA 15:1045–1058

Sachs AB, Davis RW (1989) The poly(A) binding protein is required for poly(A) shortening and 60S ribosomal subunit-dependent translation initiation. Cell 58:857–867
Sachs AB, Deardorff JA (1992) Translation initiation requires the PAB-dependent poly(A) ribonuclease in yeast. Cell 70:961–973
Sachs AB, Bond MW, Kornberg RD (1986) A single gene from yeast for both nuclear and cytoplasmic polyadenylate-binding proteins: domain structure and expression. Cell 45:827–835
Sachs AB, Davis RW, Kornberg RD (1987) A single domain of yeast poly(A)-binding protein is necessary and sufficient for RNA binding and cell viability. Mol Cell Biol 7:3268–3276
Saint-Georges Y, Garcia M, Delaveau T, Jourdren L, Le Crom S, Lemoine S, Tanty V, Devaux F, Jacq C (2008) Yeast mitochondrial biogenesis: a role for the PUF RNA-binding protein Puf3p in mRNA localization. PLoS ONE 3:e2293
Salgado-Garrido J, Bragado-Nilsson E, Kandels-Lewis S, Séraphin B (1999) Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin. EMBO J 18:3451–3462
San Paolo S, Vanacova S, Schenk L, Scherrer T, Blank D, Keller W, Gerber AP (2009) Distinct roles of non-canonical poly(A) polymerases in RNA metabolism. PLoS Genet 5:e1000555
Santiveri CM, Mirassou Y, Rico-Lastres P, Martínez-Lumbreras S, Pérez-Cañadillas JM (2011) Pub1p C-terminal RRM domain interacts with Tif4631p through a conserved region neighbouring the Pab1p binding site. PLoS ONE 6:e24481
Sayani S, Chanfreau GF (2012) Sequential RNA degradation pathways provide a fail-safe mechanism to limit the accumulation of unspliced transcripts in Saccharomyces cerevisiae. RNA 18:1563–1572
Schaeffer D, Tsanova B, Barbas A, Reis FP, Dastidar EG, Sanchez-Rotunno M, Arraiano CM, van Hoof A (2009) The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities. Nat Struct Mol Biol 16:56–62
Schmidt K, Xu Z, Mathews DH, Butler JS (2012) Air proteins control differential TRAMP substrate specificity for nuclear RNA surveillance. RNA 18:1934–1945
Schneider C, Anderson JT, Tollervey D (2007) The exosome subunit Rrp44 plays a direct role in RNA substrate recognition. Mol Cell 27:324–331
Schneider C, Leung E, Brown J, Tollervey D (2009) The N-terminal PIN domain of the exosome subunit Rrp44 harbors endonuclease activity and tethers Rrp44 to the yeast core exosome. Nucleic Acids Res 37:1127–1140
Schulz D, Schwalb B, Kiesel A, Baejen C, Torkler P, Gagneur J, Soeding J, Cramer P (2013) Transcriptome surveillance by selective termination of noncoding RNA synthesis. Cell 155:1075–1087
Schwartz D, Decker CJ, Parker R (2003) The enhancer of decapping proteins, Edc1p and Edc2p, bind RNA and stimulate the activity of the decapping enzyme. RNA 9:239–251
Shah KH, Zhang B, Ramachandran V, Herman PK (2013) Processing body and stress granule assembly occur by independent and differentially regulated pathways in Saccharomyces cerevisiae. Genetics 193:109–123
Shalem O, Dahan O, Levo M, Martinez MR, Furman I, Segal E, Pilpel Y (2008) Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. Mol Syst Biol 4:223
Sharif H, Conti E (2013) Architecture of the Lsm1-7-Pat1 complex: a conserved assembly in eukaryotic mRNA turnover. Cell Rep 5:283–291
Sharif H, Ozgur S, Sharma K, Basquin C, Urlaub H, Conti E (2013) Structural analysis of the yeast Dhh1-Pat1 complex reveals how Dhh1 engages Pat1, Edc3 and RNA in mutually exclusive interactions. Nucleic Acids Res 41:8377–8390
Shatkin AJ, Manley JL (2000) The ends of the affair: capping and polyadenylation. Nat Struct Biol 7:838–842
Sheiness D, Puckett L, Darnell JE (1975) Possible relationship of poly(A) shortening to mRNA turnover. Proc Natl Acad Sci USA 72:1077–1081
Sheth U, Parker R (2003) Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. Science 300:805–808

Sinturel F, Bréchemier-Baey D, Kiledjian M, Condon C, Bénard L (2012) Activation of 5′-3′ exoribonuclease Xrn1 by cofactor Dcs1 is essential for mitochondrial function in yeast. Proc Natl Acad Sci USA 109:8264–8269

Sripati CE, Groner Y, Warner JR (1976) Methylated, blocked 5′ termini of yeast mRNA. J Biol Chem 251:2898–2904

Stead JA, Costello JL, Livingstone MJ, Mitchell P (2007) The PMC2NT domain of the catalytic exosome subunit Rrp6p provides the interface for binding with its cofactor Rrp47p, a nucleic acid-binding protein. Nucleic Acids Res 35:5556–5567

Steiger M, Carr-Schmid A, Schwartz DC, Kiledjian M, Parker R (2003) Analysis of recombinant yeast decapping enzyme. RNA 9:231–238

Stepien PP, Margossian SP, Landsman D, Butow RA (1992) The yeast nuclear gene suv3 affecting mitochondrial post-transcriptional processes encodes a putative ATP-dependent RNA helicase. Proc Natl Acad Sci USA 89:6813–6817

Stevens A (1978) An exoribonuclease from *Saccharomyces cerevisiae*: effect of modifications of 5′ end groups on the hydrolysis of substrates to 5′ mononucleotides. Biochem Biophys Res Commun 81:656–661

Stevens A (1980) An mRNA decapping enzyme from ribosomes of *Saccharomyces cerevisiae*. Biochem Biophys Res Commun 96:1150–1155

Stevens A (1988) mRNA-decapping enzyme from *Saccharomyces cerevisiae*: purification and unique specificity for long RNA chains. Mol Cell Biol 8:2005–2010

Sun M, Schwalb B, Schulz D, Pirkl N, Etzold S, Lariviere L, Maier KC, Seizl M, Tresch A, Cramer P (2012) Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. Genome Res 22:1350–1359

Sun M, Schwalb B, Pirkl N, Maier KC, Schenk A, Failmezger H, Tresch A, Cramer P (2013) Global analysis of Eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. Mol Cell 52:52–62

Sweet T, Kovalak C, Coller J (2012) The DEAD-box protein Dhh1 promotes decapping by slowing ribosome movement. PLoS Biol 10:e1001342

Sweet TJ, Boyer B, Hu W, Baker KE, Coller J (2007) Microtubule disruption stimulates P-body formation. RNA 13:493–502

Swisher KD, Parker R (2010) Localization to, and effects of Pbp1, Pbp4, Lsm12, Dhh1, and Pab1 on stress granules in *Saccharomyces cerevisiae*. PLoS ONE 5:e10006

Synowsky SA, Heck AJR (2008) The yeast Ski complex is a hetero-tetramer. Protein Sci 17:119–125

Synowsky SA, van Wijk M, Raijmakers R, Heck AJR (2009) Comparative multiplexed mass spectrometric analyses of endogenously expressed yeast nuclear and cytoplasmic exosomes. J Mol Biol 385:1300–1313

Tenenbaum SA, Carson CC, Lager PJ, Keene JD (2000) Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. Proc Natl Acad Sci USA 97:14085–14090

Tharun S, Parker R (2001) Targeting an mRNA for decapping: displacement of translation factors and association of the Lsm1p-7p complex on deadenylated yeast mRNAs. Mol Cell 8:1075–1083

Tharun S, He W, Mayes AE, Lennertz P, Beggs JD, Parker R (2000) Yeast Sm-like proteins function in mRNA decapping and decay. Nature 404:515–518

Toh-E A, Guerry P, Wickner RB (1978) Chromosomal superkiller mutants of *Saccharomyces cerevisiae*. J Bacteriol 136:1002–1007

Trapman J, Retèl J, Planta RJ (1975) Ribosomal precursor particles from yeast. Exp Cell Res 90:95–104

Trcek T, Larson DR, Moldón A, Query CC, Singer RH (2011) Single-molecule mRNA decay measurements reveal promoter-regulated mRNA stability in yeast. Cell 147:1484–1497

Tuck AC, Tollervey D (2013) A Transcriptome-wide Atlas of RNP composition reveals diverse classes of mRNAs and lncRNAs. Cell 154:996–1009

Tucker M, Staples RR, Valencia-Sanchez MA, Muhlrad D, Parker R (2002) Ccr4p is the catalytic subunit of a Ccr4p/Pop2p/Notp mRNA deadenylase complex in *Saccharomyces cerevisiae*. EMBO J 21:1427–1436

Turk EM, Das V, Seibert RD, Andrulis ED (2013) The mitochondrial RNA landscape of Saccharomyces cerevisiae. PLoS ONE 8:e78105

Udem SA, Kaufman K, Warner JR (1971) Small ribosomal ribonucleic acid species of Saccharomyces cerevisiae. J Bacteriol 105:101–106

Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB (2003) CLIP identifies Nova-regulated RNA networks in the brain. Science 302:1212–1215

Vanácová S, Wolf J, Martin G, Blank D, Dettwiler S, Friedlein A, Langen H, Keith G, Keller W (2005) A new yeast poly(A) polymerase complex involved in RNA quality control. PLoS Biol 3:e189

Wang Z, Kiledjian M (2001) Functional link between the mammalian exosome and mRNA decapping. Cell 107:751–762

Wang L, Lewis MS, Johnson AW (2005) Domain interactions within the Ski2/3/8 complex and between the Ski complex and Ski7p. RNA 11:1291–1302

Wang SW, Toda T, MacCallum R, Harris AL, Norbury C (2000) Cid1, a fission yeast protein required for S-M checkpoint control when DNA polymerase delta or epsilon is inactivated. Mol Cell Biol 20:3234–3244

Wasmuth EV, Lima CD (2012) Exo- and endoribonucleolytic activities of yeast cytoplasmic and nuclear RNA exosomes are dependent on the noncatalytic core and central channel. Mol Cell 48:133–144

Wegierski T, Dmochowska A, Jabłonowska A, Dziembowski A, Bartnik E, Stepień PP (1998) Yeast nuclear PET127 gene can suppress deletions of the SUV3 or DSS1 genes: an indication of a functional interaction between 3´ and 5´ ends of mitochondrial mRNAs. Acta Biochim Pol 45:935–940

Wery M, Ruidant S, Schillewaert S, Lepore N, Lafontaine DLJ (2009) The nuclear poly(A) polymerase and Exosome cofactor Trf5 is recruited cotranscriptionally to nucleolar surveillance. RNA 15:406–419

Wickens M, Bernstein DS, Kimble J, Parker R (2002) A PUF family portrait: 3´UTR regulation as a way of life. Trends Genet TIG 18:150–157

Widner WR, Wickner RB (1993) Evidence that the SKI antiviral system of *Saccharomyces cerevisiae* acts by blocking expression of viral mRNA. Mol Cell Biol 13:4331–4341

Wiesenberger G, Fox TD (1997) Pet127p, a membrane-associated protein involved in stability and processing of *Saccharomyces cerevisiae* mitochondrial RNAs. Mol Cell Biol 17:2816–2824

Wlotzka W, Kudla G, Granneman S, Tollervey D (2011) The nuclear RNA polymerase II surveillance system targets polymerase III transcripts. EMBO J 30:1790–1803

Wu D, Muhlrad D, Bowler MW, Liu Z, Parker R, Song H (2013) Lsm2 and Lsm3 bridge the interaction of the Lsm1-7 complex with Pat1 for decapping activation. Cell Res 24(2):233–246

Wyers F, Rougemaille M, Badis G, Rousselle J-C, Dufour M-E, Boulay J, Régnault B, Devaux F, Namane A, Séraphin B et al (2005) Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. Cell 121:725–737

Xiang S, Cooper-Morgan A, Jiao X, Kiledjian M, Manley JL, Tong L (2009) Structure and function of the 5´–>3´ exoribonuclease Rat1 and its activating partner Rai1. Nature 458:784–788

Xue Y, Bai X, Lee I, Kallstrom G, Ho J, Brown J, Stevens A, Johnson AW (2000) *Saccharomyces cerevisiae* RAI1 (YGL246c) is homologous to human DOM3Z and encodes a protein that binds the nuclear exoribonuclease Rat1p. Mol Cell Biol 20:4006–4015

Yamasaki S, Anderson P (2008) Reprogramming mRNA translation during stress. Curr Opin Cell Biol 20:222–226

Zaman S, Lippman SI, Schneper L, Slonim N, Broach JR (2009) Glucose regulates transcription in yeast through a network of signaling pathways. Mol Syst Biol 5(245)

Chapter 8
Cytoplasmic mRNA Surveillance Pathways

Zaineb Fourati and Marc Graille

Abstract During mRNA synthesis and maturation, the introduction of errors can strongly influence the expression of certain genes and/or the activity of the proteins for which they encode. To minimise these defects, eukaryotic cells have evolved several cytoplasmic and translation-dependent quality control pathways aimed at detecting and degrading mRNAs that would lead to the production of aberrant proteins. The nonsense-mediated mRNA decay pathway (NMD) clears cells from mRNAs harbouring premature in-frame stop codons. Two other pathways (NSD for nonstop decay and NGD for No-Go decay) degrade mRNAs on which ribosomes have stalled during elongation. In this chapter, we describe the current knowledge on the biological roles and molecular mechanisms of these surveillance pathways, which were mainly unravelled using baker's yeast as model system.

Contents

Z. Fourati · M. Graille (✉)
Laboratoire de Biochimie, Ecole Polytechnique, CNRS UMR7654,
Palaiseau 91128 Cedex, France
e-mail: marc.graille@polytechnique.edu

A. Sesma and T. von der Haar (eds.), *Fungal RNA Biology*,
DOI: 10.1007/978-3-319-05687-6_8,

Introduction

In eukaryotes, the production of functional translatable mRNAs requires several maturation steps (splicing, capping, polyadenylation, export, …), all of which offer the possibility for introducing errors. The translation of such faulty mRNAs would produce aberrant proteins, which could have dramatic effects and lead to diseases or even cell death. However, these aberrant mRNAs are rarely translated as eukaryotic cells have evolved numerous surveillance (or QC for quality control) pathways dedicated to the detection and the rapid degradation of these mRNAs and to the concomitant clearance of nascent proteins derived from these mRNAs. The most extensively described process is the nonsense-mediated mRNA decay pathway (known as NMD), which clears cells from mRNAs containing in-frame premature termination codons (PTC) (Kervestin and Jacobson 2012; Losson and Lacroute 1979). Other QC pathways specialised in the rapid degradation of mRNAs responsible for translation elongation stalls have been described more recently. The nonstop (or NSD) and No-Go (or NGD) mRNA decay pathways degrade mRNAs lacking stop codons (Frischmeyer et al. 2002; van Hoof et al. 2002) or mRNAs inducing strong translational stalls (Doma and Parker 2006), respectively. These evolutionarily conserved mechanisms have been discovered and deeply characterised using *Saccharomyces cerevisiae* yeast. This chapter presents an overview of these cytoplasmic and translation-dependent mRNA decay pathways as a wealth of information obtained within the last years offers a more detailed understanding of these QC mechanisms.

The Nonsense-Mediated mRNA Decay Pathway

NMD Substrates

NMD rids cells from mRNAs harbouring a PTC and thereby prevents the accumulation of potentially harmful truncated proteins. PTC resulting in NMD activation can occur in mRNAs due to genetic mutations, transcription and/or mRNA maturation errors, especially splicing defaults (Kervestin and Jacobson 2012; Mitrovich and Anderson 2000). NMD substrates also include bicistronic mRNAs, pseudogene-derived transcripts, mRNAs subjected to leaky scanning leading to

translation initiation errors or to frameshifting, or mRNAs with upstream reading frames (uORFs) (He et al. 2003; Ruiz-Echevarria and Peltz 2000; Welch and Jacobson 1999). In mammals, they also arise from alternative splicing of mRNAs (Hansen et al. 2009) or are produced by genes undergoing programmed rearrangement such as those encoding antibodies, B and T cell receptors (Li and Wilkinson 1998). NMD is also activated by a stop codon followed by normal or biologically regulated long 3′ UTRs, which mimic a premature termination context (Kebaara and Atkin 2009; Muhlrad and Parker 1999).

Beyond mRNA QC, NMD also modulates the cellular level of up to 10 % of normal genes in *S. cerevisiae*, *D. melanogaster* and humans and hence directly regulates the expression of many physiological transcripts (He et al. 2003; Mendell et al. 2004; Rehwinkel et al. 2005; Wittmann et al. 2006).

NMD Factors

NMD is activated when the stop codon present in the ribosomal A-site is recognised as premature. This pathway relies on the NMD specific factors and in particular, the three conserved Upf proteins: Upf1, Upf2, and Upf3 initially identified in *S. cerevisiae* and *C. elegans* (Smg2, 3 and 4, respectively) (Cui et al. 1995; Leeds et al. 1992; Pulak and Anderson 1993). Mutations of *UPF* genes lead to a specific stabilisation of PTC-containing transcripts (He et al. 1997).

The Upf proteins interact together at the premature stop codon and form the surveillance UPF complex, where Upf1 is assumed to be the key effector of NMD while Upf2 and Upf3 act as essential regulators of its function.

Upf1

Upf1 is a large cytoplasmic protein composed of two functional domains: an N-terminal Cysteine- and Histidine-rich zinc-finger domain (CH domain) and a larger C-terminal helicase domain from the SF1 family (de la Cruz et al. 1999). The CH domain has a RING-box architecture and exhibits U3 ubiquitin-ligase activity that may be involved in the elimination of the aberrant peptide by the proteasome (Kadlec et al. 2006; Takahashi et al. 2008). This domain is also involved in Upf1 interaction with Rps26 from the ribosomal 40S subunit (Min et al. 2013). The helicase domain consists of two canonical RecA-like subdomains with two additional inserted subdomains (called 1B and 1C) and exhibits ATPase and RNA unwinding activities (Fig. 8.1). Both activities are essential for NMD and are downregulated by the CH domain (Bhattacharya et al. 2000; Chamieh et al. 2008; Czaplinski et al. 1995; Weng et al. 1996).

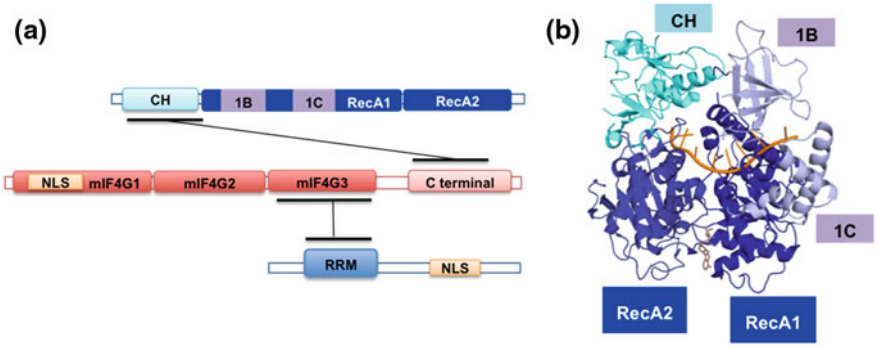


Fig. 8.1 The architecture of the Upf complex. **a** The modular organisation of Upf1 (*top*), Upf2 (*middle*) and Upf3 (*bottom*) proteins. The interacting domains of Upf1, Upf2 and Upf3 are connected by *black lines*. **b** The X-ray structure of yeast Upf1 in complex with a poly(U)$_9$ RNA (*orange* and an ATP analog (ADP-AlF$_4^-$, *brown*). The CH domain is coloured in *cyan*, the RecA1 and RecA2 domains are coloured in *dark blue* and the 1B and the 1C insertions are coloured in *light blue*

Upf3

Upf3 is a small protein with a conserved central RNA recognition motif (RRM), which is unable to bind RNA in vitro (Kadlec et al. 2004). Upf3 C-terminal domain harbours a functional Nuclear localisation signal (NLS) allowing its shuttling between the nucleus and the cytoplasm (Fig. 8.1) (Lee and Culbertson 1995; Shirley et al. 1998).

Upf2

Upf2 is the largest protein of the surveillance complex. It harbours three conserved mIF4G-like domains followed by an acidic linker and a small C-terminal domain (Fig. 8.1) (Chakrabarti et al. 2011; Clerici et al. 2009; He et al. 1997; Kadlec et al. 2006). Upf2 is generally considered as the scaffold protein within the UPF complex as it bridges Upf1 to Upf3. Indeed, Upf2 interacts with the Upf3 RRM domain through its third mIF4G domain and with the Upf1 CH domain via its C-terminal domain.

However, beyond its scaffolding role, Upf2 also enhances Upf1 enzymatic activities (Chamieh et al. 2008). Indeed, Upf2 binding to the Upf1 CH domain displaces it by 120° relative to the Upf1 helicase domain, thus releasing its cis-inhibitory effect on helicase and ATPase activities (Chakrabarti et al. 2011; Clerici et al. 2009).

The role of the two first mIF4G domains remains unclear, although their deletion abolishes NMD in yeast without affecting the formation of the UPF complex (He et al. 1997). The first mIF4G domain was proposed to harbour a putative conserved "NLS" whose deletion provokes severe NMD defects in yeast.

However, yeast Upf2 is cytoplasmic and the NMD defects observed upon "NLS" depletion are unlikely to be caused by Upf2 mislocalisation (He and Jacobson 1995). This first mIF4G domain is phosphorylated in yeast but the precise role of this post-translational modification in NMD remains unclear (Wang et al. 2006).

Other Yeast NMD Factors

Beyond the central UPF complex, other yeast proteins have been suggested to play secondary roles in NMD but their precise role is still controversial.

Hrp1 is an essential nucleocytoplasmic protein that stabilises the mRNA 3′ poly(A) tail thus contributing to the polyadenylation process (Kessler et al. 1997). Hrp1 also interacts with Upf1 and promotes NMD activation by recognising specific sequences located downstream of the PTC (called DSE for Downstream Sequence Element) (Gonzalez et al. 2000). Hrp1 was thus suggested to be a "marker" protein displaced by the translating ribosome when the stop codon is 'normal', but remaining tethered to PTC-containing mRNA, thereby triggering NMD. However, these DSE were found in several NMD mRNA reporters (PGK1, HIS4, ADE3 and GCN4) but share a weak sequence consensus, while other NMD substrates are free of DSE (Hagan et al. 1995; Ruiz-Echevarria et al. 1998; Zhang et al. 1995). Hence, this model could not be adopted as a generalised NMD activation process.

The deletion of the *EBS1* gene provokes a slight but consistent stabilisation of several NMD substrates (Luke et al. 2007). The Ebs1 protein was proposed to harbour an N-terminal 14-3-3 domain and to be a putative orthologue of human SMG7, which is involved in UPF1 dephosphorylation (Luke et al. 2007; Ohnishi et al. 2003). Although Upf1 is phosphorylated in yeast (Lasalde et al. 2013; Wang et al. 2006), no clear evidence implicates Ebs1 in the sensing of Upf1 phosphorylation status. Ebs1 may rather influence NMD by inhibiting translation (Ford et al. 2006).

The DEAD-box RNA helicase Dbp2 associates with Upf1 and is involved both in NMD and rRNA processing in yeast (Bond et al. 2001; He and Jacobson 1995). Its human orthologue, p68 (Ddx5), associates with Upf3b and activates the NMD-mediated regulation of several specific genes, including its own gene (Bond et al. 2001; He and Jacobson 1995). Hence, Dbp2 is probably the most interesting factor whose role in NMD should be clarified.

NMD Mechanism

NMD is classically considered as a three-step mechanism, where the first step consists in the recruitment of the canonical translation termination machinery (eRF1 and eRF3) upon entry of a stop codon in the ribosomal A-site. The second step is the discrimination between mRNAs harbouring PTC versus those with

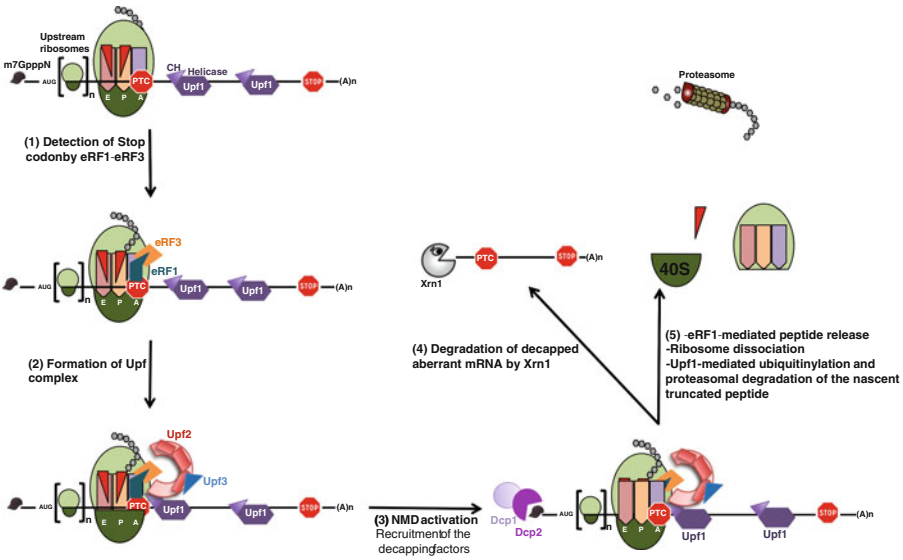


Fig. 8.2 A unified model for NMD pathway. If the recognition of a stop codon in the ribosomal A-site is accompanied by the enrichment of Upf1 on an abnormally long 3′ UTR, the Upf2 and Upf3 proteins will be recruited to form the surveillance complex, thereby signalling for the presence of a premature stop codon. This complex will enhance the degradation of the faulty mRNA as well as the proteasomal decay of the truncated peptide

normal stops. The final step consists in the rapid decay of the faulty mRNA by the cytoplasmic degradation machinery (Fig. 8.2). Accordingly, both the first and third steps involve the cellular effectors of the canonical translation termination (see Chap. 5 for details) and mRNA degradation (see Chap. 7 for details) pathways, respectively. Only the second step is carried out by the NMD specific Upf factors.

In this section, we will mainly focus on the last two steps as the recognition of stop codons has already been described in Chap. 5.

PTC Recognition and NMD Activation

Following the entry of a stop codon in the ribosomal A-site, several elements will sense the termination context and if it is detected as aberrant, will trigger a cascade of events ending in the degradation of the faulty mRNA.

It is now assumed that most of the NMD events in yeast can be explained by the 'faux 3′ UTR model', where the abnormally long sequence downstream the PTC constitutes an NMD activating signal (Amrani et al. 2004; Kervestin and Jacobson 2012). It was initially speculated that the Pab1 failure to interact with eRF3, due to the remoteness of the 3′ poly(A) tail, was the NMD triggering signal. Thus, eRF3 would bind Upf1, implying that Pab1 and Upf1 compete for eRF3 binding, thus handling the balance between normal and premature termination events

(Kervestin et al. 2012). Strong support for this hypothesis came from experiments showing that artificial tethering of yeast Pab1 in the vicinity of the PTC cause stabilisation of the corresponding mRNA (Amrani et al. 2004). In addition, artificial shortening of the 'faux 3′ UTR' by deleting the region downstream of a PTC stabilises the faulty mRNA (Hagan et al. 1995; Peltz et al. 1993). Some aspects of this model have been validated, but others have not. For instance, the requirement of the Pab1-eRF3 interaction to antagonise NMD has been discarded, as the absence of Pab1 or the deletion of the Pab1-interacting region from eRF3 do not convert a normal mRNA into an NMD substrate (Kervestin et al. 2012; Meaux et al. 2008). Rather, the key requirement for NMD activation by a 'faux 3′ UTR' context would be a proper interaction between eRF3 and Upf1. The lower efficiency of a premature termination event could indeed cause inefficient release of eRF3 thus prompting Upf1 recruitment to the PTC (Amrani et al. 2004; Kervestin and Jacobson 2012). Recent studies in human cells showed that Upf1 binds specifically to the 3′ UTR region in a length-dependent manner (Hogg and Goff 2010; Hwang et al. 2010; Kurosaki and Maquat 2013; Shigeoka et al. 2012; Zund et al. 2013). Accordingly, Upf1 is believed to sense the 3′ UTR length and to associate with long 3′ UTR-containing mRNAs, thus targeting them to NMD. This is further supported by the observation that yeast Upf1 associates preferentially with NMD substrates rather than normal mRNAs (Johansson et al. 2007). Upf1 could then recruit Upf2 and Upf3 to PTC-containing transcripts and then signal these as aberrant mRNAs to be degraded (Fig. 8.2).

This model reconciles several discrepancies between yeast and higher eukaryotes and corresponds to the most elaborated manner to interpret the differences between a normal and a 'premature' termination event. However, some twilight zones still exist and require further studies to be properly elucidated.

Faulty mRNA Degradation

The predominant mRNA decay pathway involved in yeast NMD is the 5′-3′ decay pathway (Hagan et al. 1995). Compared to normal mRNA decay, the deadenylation step is skipped in NMD and the PTC-containing transcripts undergo rapid decapping followed by subsequent exonucleolytic degradation by Xrn1 (Fig. 8.2) (Muhlrad and Parker 1994). Upf1 was proposed to recruit the decapping enzyme Dcp2 through the decapping activators Pat1 and Edc3 (He and Jacobson 2001; Swisher and Parker 2011). However, in the absence of the 5′-3′ degradation pathway, the faulty mRNA can undergo a slower 3′-5′ degradation involving Ski7 and the exosome (Mitchell and Tollervey 2003).

Proteasomal Decay of the Truncated Peptide

Beyond faulty mRNA decay, NMD also activates the rapid degradation of the truncated polypeptide by the ubiquitin-proteasome pathway (Kuroha et al. 2013;

Kuroha et al. 2009). The truncated nascent peptide will be released from the ribosome through the action of the eRF1-eRF3 translation termination factors. Upf1 seems to enhance the degradation of this truncated peptide by acting as an E3 ubiquitin ligase. Indeed, its N-terminal CH domain is structurally homologous to E3 RING finger domains and associates to the E2 enzyme Ubc3 (Takahashi et al. 2008).

NMD Factors in Higher Eukaryotes

In metazoa, NMD is a more sophisticated process and hence relies on the involvement of additional factors that are absent in yeast. The description of these factors is beyond the scope of this book, which focuses on fungi, but one can briefly mention some of these. Indeed, several SMG proteins (SMG-1 and SMG-5 to SMG-9) are involved in the regulation of UPF1 phosphorylation status. In addition, SMG-6 endonucleolytically cleaves NMD substrates in human and *D. melanogaster* (Eberle et al. 2009; Gatfield and Izaurralde 2004). Finally, the exon junction complex (EJC) is involved in the degradation of a subset of human NMD substrates (Buhler et al. 2006; Sauliere et al. 2010).

NMD Importance in Human: Involvement in Genetic Diseases and in Some Cancers

Although this chapter has almost exclusively focused on yeast NMD, this QC pathway is conserved in eukaryotes and has biological implications in human health. Indeed, it is estimated that PTC-containing mRNAs are responsible for about one third of inherited genetic disorders such as Duchenne muscular dystrophy or some forms of cystic fibrosis as well as many forms of cancer. In some instances, the truncated proteins produced by these mRNAs may be very harmful or have a dominant negative effect. In some other cases, such truncated proteins may be partially active and could, when properly expressed, decrease the disease severity. Hence, NMD is not always beneficial and could rather be seen as a double-edged sword preventing cells from producing truncated proteins that could do damage but also eliminating mRNAs encoding truncated proteins that could function normally.

It has been reported that the NMD is specifically repressed in some cancers and that this repression provokes the anarchic proliferation of the tumour cells (Gardner 2010; Wang et al. 2011a; Wang et al. 2011b). Conversely, growing evidence shows that inhibiting NMD in the tumour could play a preventive role against cancer. Indeed, NMD inhibition by siRNA (short interfering RNAs)-mediated silencing of SMG-11 or UPF1 has proved to be efficient for tumour regression by inducing an immune response against new antigens expressed in the

tumour (Gilboa 2013; Pastor et al. 2010). In the case of the genetic disorders caused by PTC, the healing strategy is rather based on inducing selective PTC read-through (for a recent review, see Bidou et al. 2012). Aminoglycosides, especially gentamicin, were first used as PTC read-through inducers to restore CFTR expression in several cases of cystic fibrosis (Bedwell et al. 1997; Wilschanski et al. 2003). However, due to their toxicity and their random efficiency, aminoglycosides are now replaced by a new molecule called PTC124, which proved to be less toxic and more efficient in inducing specific PTC read-through (Welch et al. 2007).

Quality Control Pathways Dealing with Translation Elongation Arrests

In-frame stop codons are not only crucial for correct translation termination but also for proper recycling of ribosomes and subsequent rounds of translation (see Chap. 5). Hence, mRNAs lacking stop codons or inducing strong translational stalls would trap translating ribosomes and the accumulation of these mRNAs would deplete cells from functional ribosomes. To avoid this, cells have evolved two other QC pathways (Graille and Seraphin 2012). The nonstop decay (or NSD) detects and degrades mRNAs lacking stop codons. The No-Go decay (NGD) pathway clears cells from mRNAs causing ribosomal stalls during elongation. Both pathways also trigger the degradation of the polypeptide derived from these aberrant mRNAs. These pathways have been initially identified in yeast using artificial reporters but natural substrates were later identified, rationalising their biological importance. Finally, the molecular mechanisms of these pathways have been largely deciphered very recently using yeast as a model system.

Nonstop mRNA Decay or NSD

Poly(A)$^+$ NSD Substrates

The absence of in-frame stop codon within mRNAs (hereafter named nonstop mRNAs) can arise from single-point mutants converting a stop codon into a sense codon. Besides, under some circumstances that decrease stop codon recognition efficiency, ribosomes can also perform stop codon read-through and synthesise longer proteins. These two classes of NSD substrates are poly(A)$^+$ mRNAs and it is commonly considered that ribosomes translating these mRNAs will be stalled on the 3′ poly(A) tail and produce a polylysine extension at the C-terminal extremity of these extended proteins that will remain covalently bound to the P-site tRNA.

However, such substrates are rather rare as 3′ UTRs are generally rich in in-frame stop codons.

The poly(A)$^+$ NSD mRNAs are strongly destabilised both in yeast and mammals (Frischmeyer et al. 2002; van Hoof et al. 2002) through the exo- and endonucleolytic activities of the Rrp44/Dis3 exosome catalytic subunit as well as the SKI complex and their associated factor Ski7 (a yeast-specific protein and a member of the eEF1A translational GTPase family; Schaeffer and van Hoof 2011; van Hoof et al. 2002). In human cells, the role played by Ski7 in yeast NSD is performed by Hbs1, another member of eEF1A translational GTPase family (Saito et al. 2013). Concomitant to the accelerated mRNA decay, the corresponding nonstop proteins are not detected in the cells suggesting a translational repression mechanism, a higher instability or both (Dimitrova et al. 2009; Inada and Aiba 2005; Ito-Harashima et al. 2007). The levels of nonstop proteins but not nonstop mRNAs are strongly decreased by proteins linked to the proteasome such as the Ltn1 RING-domain-type E3 ubiquitin ligase (Bengtson and Joazeiro 2010; Wilson et al. 2007). These proteins ubiquitinylate nascent nonstop proteins, further triggering their degradation by the proteasome (Bengtson and Joazeiro 2010). Altogether, the expression of nonstop poly(A)$^+$ mRNAs exhibits three levels of regulation: mRNA stability, translational repression and nonstop protein stability.

Poly(A)-less NSD Substrates

Another type of nonstop mRNAs can arise in vivo following endonucleolytic cleavage of mRNAs (i.e. in the case of NGD, see section "No-Go decay or NGD"). These are poly(A)-less mRNAs and lead to ribosomes stalled at the 3′ end of these mRNAs, unable to recycle, thereby producing a shorter polypeptide chain remaining attached to the P-site tRNA. These mRNAs are also highly unstable but several discrepancies exist when compared to the decay pathway described for poly(A)$^+$ nonstop mRNAs. In particular, while the latter requires both the N and C-terminal domains of Ski7 to be degraded, the decay of poly(A)-less nonstop mRNAs does not require the Ski7 C-terminal domain (Meaux and Van Hoof 2006). In addition, proteins derived from these nonstop poly(A)-less mRNAs are produced at a low level. This could be caused by a reduced translation of the nonstop mRNA and/or to a decreased protein stability due to defects in peptide release from the ribosome because of the lack of stop codon. Again, the Ltn1 protein, together with Cdc48 and the RQC complex, address these nonstop proteins to the proteasome for degradation (Brandman et al. 2012; Defenouillere et al. 2013). It was also shown that the protein production from poly(A)-less nonstop mRNA is dependent on the Dom34 and Hbs1 proteins (Kobayashi et al. 2010; see below for details on these two proteins).

No-Go Decay or NGD

A third class of aberrant mRNAs induce translational stalls due to the presence of a stable stem loop, pseudoknot, rare codons, stretch of consecutive identical residues (either K_{12} or R_{12}) or apurinic sites (Dimitrova et al. 2009; Doma and Parker 2006; Gandhi et al. 2008; Kuroha et al. 2010). These are degraded by the NGD pathway, which endonucleolytically cleaves these mRNAs close to the stalling site prior to their degradation by classical exonucleases (Xrn1 and the exosome). The Dom34 and Hbs1 proteins, which share significant similarity to translation termination factors eRF1 and eRF3, respectively, are important but not essential for the endonucleolytic cleavage observed in NGD (Doma and Parker 2006; Kuroha et al. 2010). As a result, the levels of nascent proteins produced by these mRNAs are lower than expected (Dimitrova et al. 2009; Kuroha et al. 2010). It has recently been proposed that the *S. cerevisiae* Asc1 protein (or RACK1), a core component of the small ribosomal 40S subunit, which binds to the exit of the mRNA channel, might stimulate translational arrest, thereby leading to nascent protein degradation by the Not4 and Ltn1 E3 ubiquitin ligase proteins (Bengtson and Joazeiro 2010; Dimitrova et al. 2009; Kuroha et al. 2010; Panasenko et al. 2006).

Dom34 and Hbs1, Central Factors of these mRNA QC Pathways

During the last years, several studies have unravelled the mechanisms of NSD and NGD. Dom34 and Hbs1 appear to play a central role in these processes.

Dom34 displays strong structural similarity with class I translation termination factor eRF1 (Graille et al. 2008; Lee et al. 2007). Indeed, Dom34 is composed of three distinct domains: N-terminal, central and C-terminal domains. These domains are spatially arranged so as to mimic a tRNA with the N-terminal and central domains corresponding to the anticodon loop and amino acyl acceptor arm of the tRNAs, as observed for eRF1. Despite this structural similarity, Dom34 and eRF1 proteins display some important differences. First, the universally conserved GGQ motif from the eRF1 central domain, which enters into the ribosomal peptidyltransferase centre to catalyse the hydrolysis of the peptidyl-tRNA bond, is absent in the Dom34 central domain, indicating that Dom34 should not induce release of the nascent peptide. Second, the Dom34 N-terminal domain is structurally radically different from eRF1 N-terminal domain. In Dom34, this domain adopts an Sm/Lsm like fold, suggesting a role in RNA binding by analogy with other Sm/Lsm domains (Wilusz and Wilusz 2005).

Hbs1 belongs to the translational GTPases family encompassing bacterial and eukaryotic elongation factors EF-Tu and eEF1A as well as the eukaryotic class II release factor eRF3 and the yeast-specific Ski7 protein (Atkinson et al. 2008). Hbs1 is mainly composed of a GTPase domain followed by two β-barrels domains

(II and III) (van den Elzen et al. 2010). GTP binding to Hbs1 is required for its biological function and for its roles in mRNA QC pathways (Carr-Schmid et al. 2002; Kobayashi et al. 2010; van den Elzen et al. 2010).

Hbs1 and Dom34 proteins interact together to form a stable complex, which structurally mimics both eRF1-eRF3 and EF-Tu-tRNA complexes (Graille et al. 2008; Kobayashi et al. 2010; Kobayashi et al. 2012; Nissen et al. 1995). The similarity with the EF-Tu–tRNA complex is further reinforced by the binding mode of Dom34–Hbs1 to the ribosomal A-site of stalled ribosomes (Becker et al. 2011).

Mechanism of NSD and NGD QC Pathways

Based on all the information gathered in vivo and in vitro within recent years, there is growing evidence indicating that NGD and NSD pathways function in a very similar manner. It is now possible to propose the following mechanism for these QC pathways that degrade mRNAs impeding translation elongation by the ribosome (Fig. 8.3).

(1) The Dom34-Hbs1 complex in its GTP form is recruited to the A-site of ribosomes stalled in translation. Contrary to the eRF1-eRF3 complex that specifically recognises a stop codon in the A-site, Dom34-Hbs1 binding to the ribosome is independent of the codon present in the ribosomal A-site (Shoemaker et al. 2010). (2) The mRNA associated with the stalled ribosome is endonucleolytically cleaved mainly upstream of the stalled ribosome (Tsuboi et al. 2012). It is noteworthy that following this cleavage, when the ribosomes located upstream of the cleavage site will reach the 3′ end of the truncated mRNA, no stop codon will be present and hence, at this point NGD and NSD meet together. (3) After cleavage, the ribosome stimulates Hbs1 GTPase activity, which could be accompanied by a large conformational change of the intrinsically flexible Dom34 central domain. Dom34 will then adopt a conformation similar to the tRNA "A/A" state observed for EF-Tu-tRNA bound to the bacterial ribosome (Becker et al. 2012; Schmeing et al. 2009), with its central domain oriented towards the peptidyltransferase centre. GTP hydrolysis could also induce a rearrangement of the Hbs1 GTPase domain relative to domains II and III to adopt a conformation similar to that of the *S. pombe* Dom34-Hbs1 complex and hence lead to Hbs1 dissociation from the ribosome (Chen et al. 2010). (4) The highly conserved and essential Rli1 protein (a member of the ABC family known as ABCE1 in human) is recruited to the ribosome and binds to the same sites as Hbs1 both on the ribosome and on Dom34 (Becker et al. 2012). (5) ATP hydrolysis by Rli1 will result in ribosome dissociation (Pisarev et al. 2010) followed by mRNA degradation by the Xrn1 exonuclease and the exosome (Doma and Parker 2006). (6) The peptidyl-tRNA bound to the P-site should be released from the ribosome. In the case of ribosomes stalled after a few rounds of elongation, the nascent peptide attached to the P-site tRNA should be short and the peptidyl-tRNA could drop-off easily from the

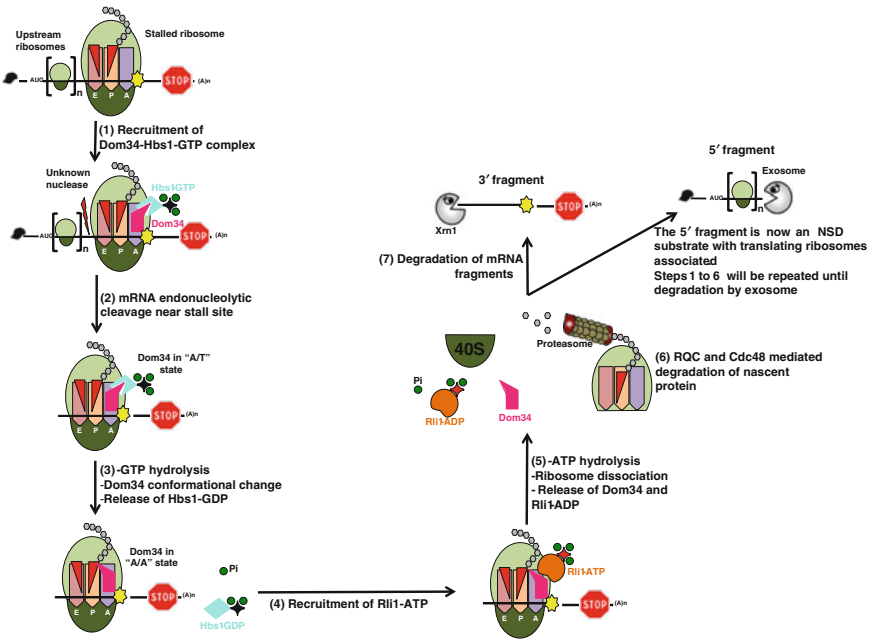


Fig. 8.3 Model for the No-Go decay molecular mechanism. Based on our current knowledge of these pathways, this model holds true for NSD pathway

ribosome. However, in the case of ribosomes stopped after several rounds of elongation, the nascent peptide will be longer, already deeply engaged into the ribosomal exit tunnel. Despite the structural similarity between Dom34 and eRF1, Dom34 does not catalyse the release of the newly synthesised protein. It has recently been shown that the RQC complex (formed by the Ltn1, Tae2 and Rqc1 proteins) together with the Cdc48 AAA^+ ATPase and its cofactors (Npl4 and Ufd1) and Not4 (in some cases) address the nascent proteins derived from NGD and NSD substrates to the proteasome for degradation (Bengtson and Joazeiro 2010; Brandman et al. 2012; Defenouillere et al. 2013; Dimitrova et al. 2009). (7) The 5′ and 3′ fragments from the defective mRNA should now be eliminated. The 5′–>3′ exonuclease Xrn1 will degrade the 3′ fragment, which does not contain a cap structure at its 5′ extremity. The 5′ fragment still contains ribosomes engaged in translation. Since the stalling site has been removed, one can imagine that translation by these ribosomes should be resumed until they reach the 3′ end of this mRNA fragment, which does not contain stop codon. This 5′ fragment then becomes an NSD substrate and ribosomes stalled at the 3′ end of this fragment should be removed by reiteration of steps 1 to 6. If there is no ribosome left on this mRNA fragment, this one can then be degraded through the action of the exosome.

Biological Implications of NSD and NGD QC Pathways

Although mainly studied in budding yeast, the NSD and NGD pathways are evolutionarily conserved, supporting that they play important biological functions. Indeed, these two QC pathways have been described for *D. melanogaster* and human cells, where they also involve the Dom34 and Hbs1 proteins (Frischmeyer et al. 2002; Passos et al. 2009; Saito et al. 2013).

The NSD pathway relies on the absence of in-frame stop codons. Mutations of the stop codon into a sense codon thereby resulting in the absence of in-frame stop codons have been documented to be responsible for two human diseases: 2, 8-dihydroxyadenine urolithiasis and hypogonadotrophic hypogonadism (Seminara et al. 2003; Taniguchi et al. 1998). In both cases, the levels of nonstop mRNA and the resulting protein are significantly reduced, suggesting that these nonstop mRNAs are cleared from cells by the NSD pathway. However, mutations in normal termination codons or stop codon read-through would not routinely initiate NSD due to the frequent occurrence of in-frame stop codons in the 3′ UTR but would rather result in C-terminally extended proteins. Hence, the evolutionary pressure that has resulted in maintenance of NSD eukaryotes should result from the presence of a non-negligible number of endogenous mRNA NSD substrates. In particular, eukaryotic genes can contain consensus sequences for 3′ end processing (i.e. cleavage and polyadenylation) within their coding region. This is the case for approximately 0.7–0.8 % of yeast (such as CBP1 and RNA14) and human genes (Frischmeyer et al. 2002; Mayer and Dieckmann 1991; Sparks and Dieckmann 1998). Prematurely polyadenylated truncated forms of the yeast CBP1 and chicken Growth hormone receptor (for GHR) mRNAs are indeed NSD substrates. Hence, the NSD pathway could, under certain physiological conditions, regulate the abundance of some mRNAs.

NGD substrates are probably more frequent than NSD substrates in cells as various events can cause translational stalls. For instance, S-adenosyl-L-Methionine rules the stability of the *A. thaliana* CGS1 mRNA encoding cystathione γ-synthase by inducing translation elongation arrest followed by mRNA endonucleolytic cleavage (Onouchi et al. 2005). Furthermore, bioinformatics searches for yeast genes containing signals susceptible to enforce ribosome pausing (i.e. stretches of at least 10 consecutive basic residues, stable stem loops or pseudo-knots, …) have identified potential NGD substrates (Dimitrova et al. 2009; Jacobs et al. 2007). Some of these were experimentally characterised. The *JJJ1*, *MAP2* and *RMP1* mRNAs induce translational arrest and release of nonstop protein products. Similarly, upon *DOM34* deletion, the steady state levels of mRNA encoding for Est2 and Bub3 are strongly stabilised while mRNA encoding for Spr6 is stabilised by two-fold (Belew et al. 2010).

The Dom34 and Hbs1 proteins have also been involved in the degradation of mRNAs containing apurinic sites, which can cause elongation stalls due to imperfect mRNA codon–tRNA anticodon base pairing (Gandhi et al. 2008). The occurrence of apurinic sites caused by chemical compounds is well characterised

in DNA as well as the associated repair mechanisms such as base excision repair (BER), which allow regenerating an intact copy of the genetic information (Robertson et al. 2009). The chemical damages underwent by RNAs are much less characterised but growing evidences suggests that mRNAs as well as non-coding RNAs can be oxidised, alkylated and damaged by other means (reviewed in Wurtmann and Wolin 2009). For instance, the oxidation of mRNAs has been shown to lead to translation elongation stalls and reduction of the production of the corresponding proteins (Shan et al. 2007). This could be due to the action of NGD and NSD pathways. Compared to damaged DNA molecules, which have to be repaired to reduce spreading of errors during cell division, one can imagine that degradation of damaged mRNAs is less energy-consuming for cells than repair, in particular for transient molecules with short half-lives such as mRNAs. Hence, the NGD pathway may be one of the mechanisms used by eukaryotic cells to degrade subsets of damaged mRNAs enforcing ribosomes to stall during translation elongation.

Finally, several observations suggest that a biologically relevant function of the Dom34-Hbs1 complex is probably related to the degradation of immature or non-functional, small ribosomal subunits that cannot elongate properly (Cole et al. 2009; LaRiviere et al. 2006; Soudet et al. 2010; Strunk et al. 2012). Hence, the Dom34-Hbs1 complex could play a predominant role by detecting and inducing the degradation of non-functional ribosomes that have passed successfully through all the check points, are able to initiate translation but are unable to proceed in elongation.

Conclusion

Baker's yeast undeniably played a central and key role for the identification and the description of these conserved translation-dependent eukaryotic mRNA QC pathways. Some twilight zones still persist and although studies performed with human cells become accessible to more laboratories, yeast will undoubtedly continue to play a major role in the future description of the still unknown steps of these processes. Among these *terra incognita* to be explored, the exact NMD mechanism responsible for the discrimination between premature and normal termination events remains to be clarified. Similarly, the molecular connection between the Upf factors and the decapping machinery remains fuzzy. The potential implication in NMD of other yet unidentified factors has also to be addressed. Regarding NSD and NGD, the description of the molecular mechanisms of these two processes has been very successful within the last 5 years and these studies also raised the veil on the mechanism of termination of the translation. However, further studies are clearly needed to decipher the physiological roles of these eukaryotic pathways. Finally, there is growing evidence linking these mRNA QC pathways with ribosome biogenesis process. Indeed, the Dbp2 rRNA processing factor is involved in NMD (Bond et al. 2001; Geissler et al. 2013), while Dom34

has been shown to play a role in a late GC checkpoint during 40S maturation (Soudet et al. 2010; Strunk et al. 2012). The relationships between these processes will have to be addressed in the future.

Acknowledgments MG acknowledges funding from the Centre National pour la Recherche Scientifique ATIP-AVENIR program, the Agence Nationale pour la Recherche (grants ANR-06-BLAN-0075-02 and ANR-11-BSV800902), the Association Française contre les Myopathies, and the Human Frontiers Science Program (grant RGP0018). ZF acknowledges the financial support from the Fondation pour la Recherche Médicale (FRM). The authors apologise for the many studies that could not be cited due to space constraints.

References

Amrani N, Ganesan R, Kervestin S, Mangus DA, Ghosh S, Jacobson A (2004) A faux 3′-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. Nature 432(7013):112–118

Atkinson GC, Baldauf SL, Hauryliuk V (2008) Evolution of nonstop, no-go and nonsense-mediated mRNA decay and their termination factor-derived components. BMC Evol Biol 8:290

Becker T, Armache JP, Jarasch A, Anger AM, Villa E, Sieber H, Motaal BA, Mielke T, Berninghausen O, Beckmann R (2011) Structure of the no-go mRNA decay complex Dom34-Hbs1 bound to a stalled 80S ribosome. Nat Struct Mol Biol 18(6):715–720

Becker T, Franckenberg S, Wickles S, Shoemaker CJ, Anger AM, Armache JP, Sieber H, Ungewickell C, Berninghausen O, Daberkow I, Karcher A, Thomm M, Hopfner KP, Green R, Beckmann R (2012) Structural basis of highly conserved ribosome recycling in eukaryotes and archaea. Nature 482(7386):501–506

Bedwell DM, Kaenjak A, Benos DJ, Bebok Z, Bubien JK, Hong J, Tousson A, Clancy JP, Sorscher EJ (1997) Suppression of a CFTR premature stop mutation in a bronchial epithelial cell line. Nat Med 3(11):1280–1284

Belew AT, Advani VM, Dinman JD (2010) Endogenous ribosomal frameshift signals operate as mRNA destabilizing elements through at least two molecular pathways in yeast. Nucleic Acids Res 39(7):2799–2808

Bengtson MH, Joazeiro CA (2010) Role of a ribosome-associated E3 ubiquitin ligase in protein quality control. Nature 467(7314):470–473

Bhattacharya A, Czaplinski K, Trifillis P, He F, Jacobson A, Peltz SW (2000) Characterization of the biochemical properties of the human Upf1 gene product that is involved in nonsense-mediated mRNA decay. RNA 6(9):1226–1235

Bidou L, Allamand V, Rousset JP, Namy O (2012) Sense from nonsense: therapies for premature stop codon diseases. Trends Mol Med 18(11):679–688

Bond AT, Mangus DA, He F, Jacobson A (2001) Absence of Dbp2p alters both nonsense-mediated mRNA decay and rRNA processing. Mol Cell Biol 21(21):7366–7379

Brandman O, Stewart-Ornstein J, Wong D, Larson A, Williams CC, Li GW, Zhou S, King D, Shen PS, Weibezahn J, Dunn JG, Rouskin S, Inada T, Frost A, Weissman JS (2012) A ribosome-bound quality control complex triggers degradation of nascent peptides and signals translation stress. Cell 151(5):1042–1054

Buhler M, Steiner S, Mohn F, Paillusson A, Muhlemann O (2006) EJC-independent degradation of nonsense immunoglobulin-mu mRNA depends on 3′ UTR length. Nat Struct Mol Biol 13(5):462–464

Carr-Schmid A, Pfund C, Craig EA, Kinzy TG (2002) Novel G-protein complex whose requirement is linked to the translational status of the cell. Mol Cell Biol 22(8):2564–2574

Chakrabarti S, Jayachandran U, Bonneau F, Fiorini F, Basquin C, Domcke S, Le Hir H, Conti E (2011) Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. Mol Cell 41(6):693–703

Chamieh H, Ballut L, Bonneau F, Le Hir H (2008) NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. Nat Struct Mol Biol 15(1):85–93

Chen L, Muhlrad D, Hauryliuk V, Cheng Z, Lim MK, Shyp V, Parker R, Song H (2010) Structure of the Dom34-Hbs1 complex and implications for no-go decay. Nat Struct Mol Biol 17(10):1233–1240

Clerici M, Mourao A, Gutsche I, Gehring NH, Hentze MW, Kulozik A, Kadlec J, Sattler M, Cusack S (2009) Unusual bipartite mode of interaction between the nonsense-mediated decay factors, UPF1 and UPF2. EMBO J 28(15):2293–2306

Cole SE, LaRiviere FJ, Merrikh CN, Moore MJ (2009) A convergence of rRNA and mRNA quality control pathways revealed by mechanistic analysis of nonfunctional rRNA decay. Mol Cell 34(4):440–450

Cui Y, Hagan KW, Zhang S, Peltz SW (1995) Identification and characterization of genes that are required for the accelerated degradation of mRNAs containing a premature translational termination codon. Genes Dev 9(4):423–436

Czaplinski K, Weng Y, Hagan KW, Peltz SW (1995) Purification and characterization of the Upf1 protein: a factor involved in translation and mRNA degradation. RNA 1(6):610–623

de la Cruz J, Kressler D, Linder P (1999) Unwinding RNA in Saccharomyces cerevisiae: DEAD-box proteins and related families. Trends Biochem Sci 24(5):192–198

Defenouillere Q, Yao Y, Mouaikel J, Namane A, Galopier A, Decourty L, Doyen A, Malabat C, Saveanu C, Jacquier A, Fromont-Racine M (2013) Cdc48-associated complex bound to 60S particles is required for the clearance of aberrant translation products. Proc Natl Acad Sci USA 110(13):5046–5051

Dimitrova LN, Kuroha K, Tatematsu T, Inada T (2009) Nascent peptide-dependent translation arrest leads to Not4p-mediated protein degradation by the proteasome. J Biol Chem 284(16):10343–10352

Doma MK, Parker R (2006) Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. Nature 440(7083):561–564

Eberle AB, Lykke-Andersen S, Muhlemann O, Jensen TH (2009) SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. Nat Struct Mol Biol 16(1):49–55

Ford AS, Guan Q, Neeno-Eckwall E, Culbertson MR (2006) Ebs1p, a negative regulator of gene expression controlled by the Upf proteins in the yeast Saccharomyces cerevisiae. Eukaryot Cell 5(2):301–312

Frischmeyer PA, van Hoof A, O'Donnell K, Guerrerio AL, Parker R, Dietz HC (2002) An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. Science 295(5563):2258–2261

Gandhi R, Manzoor M, Hudak KA (2008) Depurination of Brome mosaic virus RNA3 in vivo results in translation-dependent accelerated degradation of the viral RNA. J Biol Chem 283(47):32218–32228

Gardner LB (2010) Nonsense-mediated RNA decay regulation by cellular stress: implications for tumorigenesis. Mol Cancer Res 8(3):295–308

Gatfield D, Izaurralde E (2004) Nonsense-mediated messenger RNA decay is initiated by endonucleolytic cleavage in Drosophila. Nature 429(6991):575–578

Geissler V, Altmeyer S, Stein B, Uhlmann-Schiffler H, Stahl H (2013) The RNA helicase Ddx5/p68 binds to hUpf3 and enhances NMD of Ddx17/p72 and Smg5 mRNA. Nucleic Acids Res 41(16):7875–7888

Gilboa E (2013) Expression of new antigens on tumor cells by inhibiting nonsense-mediated mRNA decay. Immunol Res 57:44–51

Gonzalez CI, Ruiz-Echevarria MJ, Vasudevan S, Henry MF, Peltz SW (2000) The yeast hnRNP-like protein Hrp1/Nab4 marks a transcript for nonsense-mediated mRNA decay. Mol Cell 5(3):489–499

Graille M, Chaillet M, van Tilbeurgh H (2008) Structure of yeast Dom34: a protein related to translation termination factor Erf1 and involved in No-Go decay. J Biol Chem 283(11):7145–7154

Graille M, Seraphin B (2012) Surveillance pathways rescuing eukaryotic ribosomes lost in translation. Nat Rev Mol Cell Biol 13(11):727–735

Hagan KW, Ruiz-Echevarria MJ, Quan Y, Peltz SW (1995) Characterization of cis-acting sequences and decay intermediates involved in nonsense-mediated mRNA turnover. Mol Cell Biol 15(2):809–823

Hansen KD, Lareau LF, Blanchette M, Green RE, Meng Q, Rehwinkel J, Gallusser FL, Izaurralde E, Rio DC, Dudoit S, Brenner SE (2009) Genome-wide identification of alternative splice forms down-regulated by nonsense-mediated mRNA decay in Drosophila. PLoS Genet 5(6):e1000525

He F, Brown AH, Jacobson A (1997) Upf1p, Nmd2p, and Upf3p are interacting components of the yeast nonsense-mediated mRNA decay pathway. Mol Cell Biol 17(3):1580–1594

He F, Jacobson A (1995) Identification of a novel component of the nonsense-mediated mRNA decay pathway by use of an interacting protein screen. Genes Dev 9(4):437–454

He F, Jacobson A (2001) Upf1p, Nmd2p, and Upf3p regulate the decapping and exonucleolytic degradation of both nonsense-containing mRNAs and wild-type mRNAs. Mol Cell Biol 21(5):1515–1530

He F, Li X, Spatrick P, Casillo R, Dong S, Jacobson A (2003) Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5′ to 3′ mRNA decay pathways in yeast. Mol Cell 12(6):1439–1452

Hogg JR, Goff SP (2010) Upf1 senses 3′ UTR length to potentiate mRNA decay. Cell 143(3):379–389

Hwang J, Sato H, Tang Y, Matsuda D, Maquat LE (2010) UPF1 association with the cap-binding protein, CBP80, promotes nonsense-mediated mRNA decay at two distinct steps. Mol Cell 39(3):396–409

Inada T, Aiba H (2005) Translation of aberrant mRNAs lacking a termination codon or with a shortened 3′-UTR is repressed after initiation in yeast. EMBO J 24(8):1584–1595

Ito-Harashima S, Kuroha K, Tatematsu T, Inada T (2007) Translation of the poly(A) tail plays crucial roles in nonstop mRNA surveillance via translation repression and protein destabilization by proteasome in yeast. Genes Dev 21(5):519–524

Jacobs JL, Belew AT, Rakauskaite R, Dinman JD (2007) Identification of functional, endogenous programmed −1 ribosomal frameshift signals in the genome of Saccharomyces cerevisiae. Nucleic Acids Res 35(1):165–174

Johansson MJ, He F, Spatrick P, Li C, Jacobson A (2007) Association of yeast Upf1p with direct substrates of the NMD pathway. Proc Natl Acad Sci USA 104(52):20872–20877

Kadlec J, Guilligay D, Ravelli RB, Cusack S (2006) Crystal structure of the UPF2-interacting domain of nonsense-mediated mRNA decay factor UPF1. RNA 12(10):1817–1824

Kadlec J, Izaurralde E, Cusack S (2004) The structural basis for the interaction between nonsense-mediated mRNA decay factors UPF2 and UPF3. Nat Struct Mol Biol 11(4):330–337

Kebaara BW, Atkin AL (2009) Long 3′-UTRs target wild-type mRNAs for nonsense-mediated mRNA decay in Saccharomyces cerevisiae. Nucleic Acids Res 37(9):2771–2778

Kervestin S, Jacobson A (2012) NMD: a multifaceted response to premature translational termination. Nat Rev Mol Cell Biol 13(11):700–712

Kervestin S, Li C, Buckingham R, Jacobson A (2012) Testing the faux-UTR model for NMD: analysis of Upf1p and Pab1p competition for binding to eRF3/Sup35p. Biochimie 94(7):1560–1571

Kessler MM, Henry MF, Shen E, Zhao J, Gross S, Silver PA, Moore CL (1997) Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3′-end formation in yeast. Genes Dev 11(19):2545–2556

Kobayashi K, Kikuno I, Kuroha K, Saito K, Ito K, Ishitani R, Inada T, Nureki O (2010) Structural basis for mRNA surveillance by archaeal Pelota and GTP-bound EF1alpha complex. Proc Natl Acad Sci USA 107(41):17575–17579

Kobayashi K, Saito K, Ishitani R, Ito K, Nureki O (2012) Structural basis for translation termination by archaeal RF1 and GTP-bound EF1a complex. Nucleic Acids Res 40(18):9319–9328

Kuroha K, Akamatsu M, Dimitrova L, Ito T, Kato Y, Shirahige K, Inada T (2010) Receptor for activated C kinase 1 stimulates nascent polypeptide-dependent translation arrest. EMBO Rep 11(12):956–961

Kuroha K, Ando K, Nakagawa R, Inada T (2013) The Upf Factor Complex Interacts with Aberrant Products Derived from mRNAs Containing a Premature Termination Codon and Facilitates Their Proteasomal Degradation. J Biol Chem 288(40):28630–28640

Kuroha K, Tatematsu T, Inada T (2009) Upf1 stimulates degradation of the product derived from aberrant messenger RNA containing a specific nonsense mutation by the proteasome. EMBO Rep 10(11):1265–1271

Kurosaki T, Maquat LE (2013) Rules that govern UPF1 binding to mRNA 3′ UTRs. Proc Natl Acad Sci USA 110(9):3357–3362

LaRiviere FJ, Cole SE, Ferullo DJ, Moore MJ (2006) A late-acting quality control process for mature eukaryotic rRNAs. Mol Cell 24(4):619–626

Lasalde C, Rivera AV, Leon AJ, Gonzalez-Feliciano JA, Estrella LA, Rodriguez-Cruz EN, Correa ME, Cajigas IJ, Bracho DP, Vega IE, Wilkinson MF, Gonzalez CI (2014) Identification and functional analysis of novel phosphorylation sites in the RNA surveillance protein Upf1. Nucleic Acids Res 42(3):1916–1929

Lee BS, Culbertson MR (1995) Identification of an additional gene required for eukaryotic nonsense mRNA turnover. Proc Natl Acad Sci USA 92(22):10354–10358

Lee HH, Kim YS, Kim KH, Heo I, Kim SK, Kim O, Kim HK, Yoon JY, Kim HS, Kim do J, Lee SJ, Yoon HJ, Kim SJ, Lee BG, Song HK, Kim VN, Park CM, Suh SW (2007) Structural and functional insights into Dom34, a key component of No-Go mRNA decay. Mol Cell 27(6):938–950

Leeds P, Wood JM, Lee BS, Culbertson MR (1992) Gene products that promote mRNA turnover in Saccharomyces cerevisiae. Mol Cell Biol 12(5):2165–2177

Li S, Wilkinson MF (1998) Nonsense surveillance in lymphocytes? Immunity 8(2):135–141

Losson R, Lacroute F (1979) Interference of nonsense mutations with eukaryotic messenger RNA stability. Proc Natl Acad Sci USA 76(10):5134–5137

Luke B, Azzalin CM, Hug N, Deplazes A, Peter M, Lingner J (2007) Saccharomyces cerevisiae Ebs1p is a putative ortholog of human Smg7 and promotes nonsense-mediated mRNA decay. Nucleic Acids Res 35(22):7688–7697

Mayer SA, Dieckmann CL (1991) Yeast CBP1 mRNA 3′ end formation is regulated during the induction of mitochondrial function. Mol Cell Biol 11(2):813–821

Meaux S, Van Hoof A (2006) Yeast transcripts cleaved by an internal ribozyme provide new insight into the role of the cap and poly(A) tail in translation and mRNA decay. RNA 12(7):1323–1337

Meaux S, van Hoof A, Baker KE (2008) Nonsense-mediated mRNA decay in yeast does not require PAB1 or a poly(A) tail. Mol Cell 29(1):134–140

Mendell JT, Sharifi NA, Meyers JL, Martinez-Murillo F, Dietz HC (2004) Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. Nat Genet 36(10):1073–1078

Min EE, Roy B, Amrani N, He F, Jacobson A (2013) Yeast Upf1 CH domain interacts with Rps26 of the 40S ribosomal subunit. RNA 19(8):1105–1115

Mitchell P, Tollervey D (2003) An NMD pathway in yeast involving accelerated deadenylation and exosome-mediated 3′ –> 5′ degradation. Mol Cell 11(5):1405–1413

Mitrovich QM, Anderson P (2000) Unproductively spliced ribosomal protein mRNAs are natural targets of mRNA surveillance in C. elegans. Genes Dev 14(17):2173–2184

Muhlrad D, Parker R (1994) Premature translational termination triggers mRNA decapping. Nature 370(6490):578–581

Muhlrad D, Parker R (1999) Aberrant mRNAs with extended 3′ UTRs are substrates for rapid degradation by mRNA surveillance. RNA 5(10):1299–1307

Nissen P, Kjeldgaard M, Thirup S, Polekhina G, Reshetnikova L, Clark BF, Nyborg J (1995) Crystal structure of the ternary complex of Phe-tRNAPhe, EF-Tu, and a GTP analog. Science 270(5241):1464–1472

Ohnishi T, Yamashita A, Kashima I, Schell T, Anders KR, Grimson A, Hachiya T, Hentze MW, Anderson P, Ohno S (2003) Phosphorylation of hUPF1 induces formation of mRNA surveillance complexes containing hSMG-5 and hSMG-7. Mol Cell 12(5):1187–1200

Onouchi H, Nagami Y, Haraguchi Y, Nakamoto M, Nishimura Y, Sakurai R, Nagao N, Kawasaki D, Kadokura Y, Naito S (2005) Nascent peptide-mediated translation elongation arrest coupled with mRNA degradation in the CGS1 gene of Arabidopsis. Genes Dev 19(15):1799–1810

Panasenko O, Landrieux E, Feuermann M, Finka A, Paquet N, Collart MA (2006) The yeast Ccr4-Not complex controls ubiquitination of the nascent-associated polypeptide (NAC-EGD) complex. J Biol Chem 281(42):31389–31398

Passos DO, Doma MK, Shoemaker CJ, Muhlrad D, Green R, Weissman J, Hollien J, Parker R (2009) Analysis of Dom34 and Its Function in No-Go Decay. Mol Biol Cell 20(13):3025–3032

Pastor F, Kolonias D, Giangrande PH, Gilboa E (2010) Induction of tumour immunity by targeted inhibition of nonsense-mediated mRNA decay. Nature 465(7295):227–230

Peltz SW, Brown AH, Jacobson A (1993) mRNA destabilization triggered by premature translational termination depends on at least three cis-acting sequence elements and one trans-acting factor. Genes Dev 7(9):1737–1754

Pisarev AV, Skabkin MA, Pisareva VP, Skabkina OV, Rakotondrafara AM, Hentze MW, Hellen CU, Pestova TV (2010) The role of ABCE1 in eukaryotic posttermination ribosomal recycling. Mol Cell 37(2):196–210

Pulak R, Anderson P (1993) mRNA surveillance by the Caenorhabditis elegans smg genes. Genes Dev 7(10):1885–1897

Rehwinkel J, Letunic I, Raes J, Bork P, Izaurralde E (2005) Nonsense-mediated mRNA decay factors act in concert to regulate common mRNA targets. RNA 11(10):1530–1544

Robertson AB, Klungland A, Rognes T, Leiros I (2009) DNA repair in mammalian cells: Base excision repair: the long and short of it. Cell Mol Life Sci 66(6):981–993

Ruiz-Echevarria MJ, Gonzalez CI, Peltz SW (1998) Identifying the right stop: determining how the surveillance complex recognizes and degrades an aberrant mRNA. EMBO J 17(2):575–589

Ruiz-Echevarria MJ, Peltz SW (2000) The RNA binding protein Pub1 modulates the stability of transcripts containing upstream open reading frames. Cell 101(7):741–751

Saito S, Hosoda N, Hoshino S (2013) The Hbs1-Dom34 protein complex functions in non-stop mRNA decay in mammalian cells. J Biol Chem 288(24):17832–17843

Sauliere J, Haque N, Harms S, Barbosa I, Blanchette M, Le Hir H (2010) The exon junction complex differentially marks spliced junctions. Nat Struct Mol Biol 17(10):1269–1271

Schaeffer D, van Hoof A (2011) Different nuclease requirements for exosome-mediated degradation of normal and nonstop mRNAs. Proc Natl Acad Sci USA 108(6):2366–2371

Schmeing TM, Voorhees RM, Kelley AC, Gao YG, FVt Murphy, Weir JR, Ramakrishnan V (2009) The crystal structure of the ribosome bound to EF-Tu and aminoacyl-tRNA. Science 326(5953):688–694

Seminara SB, Messager S, Chatzidaki EE, Thresher RR, Acierno JS Jr, Shagoury JK, Bo-Abbas Y, Kuohung W, Schwinof KM, Hendrick AG, Zahn D, Dixon J, Kaiser UB, Slaugenhaupt SA, Gusella JF, O'Rahilly S, Carlton MB, Crowley WF Jr, Aparicio SA, Colledge WH (2003) The GPR54 gene as a regulator of puberty. N Engl J Med 349(17):1614–1627

Shan X, Chang Y, Lin CL (2007) Messenger RNA oxidation is an early event preceding cell death and causes reduced protein expression. FASEB J 21(11):2753–2764

Shigeoka T, Kato S, Kawaichi M, Ishida Y (2012) Evidence that the Upf1-related molecular motor scans the 3′-UTR to ensure mRNA integrity. Nucleic Acids Res 40(14):6887–6897

Shirley RL, Lelivelt MJ, Schenkman LR, Dahlseid JN, Culbertson MR (1998) A factor required for nonsense-mediated mRNA decay in yeast is exported from the nucleus to the cytoplasm by a nuclear export signal sequence. J Cell Sci 111(Pt 21):3129–3143

Shoemaker CJ, Eyler DE, Green R (2010) Dom34:Hbs1 promotes subunit dissociation and peptidyl-tRNA drop-off to initiate no-go decay. Science 330(6002):369–372

Soudet J, Gelugne JP, Belhabich-Baumas K, Caizergues-Ferrer M, Mougin A (2010) Immature small ribosomal subunits can engage in translation initiation in Saccharomyces cerevisiae. EMBO J 29(1):80–92

Sparks KA, Dieckmann CL (1998) Regulation of poly(A) site choice of several yeast mRNAs. Nucleic Acids Res 26(20):4676–4687

Strunk BS, Novak MN, Young CL, Karbstein K (2012) A Translation-Like Cycle Is a Quality Control Checkpoint for Maturing 40S Ribosome Subunits. Cell 150(1):111–121

Swisher KD, Parker R (2011) Interactions between Upf1 and the decapping factors Edc3 and Pat1 in Saccharomyces cerevisiae. PLoS ONE 6(10):e26547

Takahashi S, Araki Y, Ohya Y, Sakuno T, Hoshino S, Kontani K, Nishina H, Katada T (2008) Upf1 potentially serves as a RING-related E3 ubiquitin ligase via its association with Upf3 in yeast. RNA 14(9):1950–1958

Taniguchi A, Hakoda M, Yamanaka H, Terai C, Hikiji K, Kawaguchi R, Konishi N, Kashiwazaki S, Kamatani N (1998) A germline mutation abolishing the original stop codon of the human adenine phosphoribosyltransferase (APRT) gene leads to complete loss of the enzyme protein. Hum Genet 102(2):197–202

Tsuboi T, Kuroha K, Kudo K, Makino S, Inoue E, Kashima I, Inada T (2012) Dom34:hbs1 plays a general role in quality-control systems by dissociation of a stalled ribosome at the 3′ end of aberrant mRNA. Mol Cell 46(4):518–529

van den Elzen AM, Henri J, Lazar N, Gas ME, Durand D, Lacroute F, Nicaise M, van Tilbeurgh H, Seraphin B, Graille M (2010) Dissection of Dom34-Hbs1 reveals independent functions in two RNA quality control pathways. Nat Struct Mol Biol 17(12):1446–1452

van Hoof A, Frischmeyer PA, Dietz HC, Parker R (2002) Exosome-mediated recognition and degradation of mRNAs lacking a termination codon. Science 295(5563):2262–2264

Wang D, Wengrod J, Gardner LB (2011a) Overexpression of the c-myc oncogene inhibits nonsense-mediated RNA decay in B lymphocytes. J Biol Chem 286(46):40038–40043

Wang D, Zavadil J, Martin L, Parisi F, Friedman E, Levy D, Harding H, Ron D, Gardner LB (2011b) Inhibition of nonsense-mediated RNA decay by the tumor microenvironment promotes tumorigenesis. Mol Cell Biol 31(17):3670–3680

Wang W, Cajigas IJ, Peltz SW, Wilkinson MF, González CI (2006) Role for Upf2p phosphorylation in Saccharomyces cerevisiae nonsense-mediated mRNA decay. Mol Cell Biol 26(9):3390–3400

Welch EM, Barton ER, Zhuo J, Tomizawa Y, Friesen WJ, Trifillis P, Paushkin S, Patel M, Trotta CR, Hwang S, Wilde RG, Karp G, Takasugi J, Chen G, Jones S, Ren H, Moon YC, Corson D, Turpoff AA, Campbell JA, Conn MM, Khan A, Almstead NG, Hedrick J, Mollin A, Risher N, Weetall M, Yeh S, Branstrom AA, Colacino JM, Babiak J, Ju WD, Hirawat S, Northcutt VJ, Miller LL, Spatrick P, He F, Kawana M, Feng H, Jacobson A, Peltz SW, Sweeney HL (2007) PTC124 targets genetic disorders caused by nonsense mutations. Nature 447(7140):87–91

Welch EM, Jacobson A (1999) An internal open reading frame triggers nonsense-mediated decay of the yeast SPT10 mRNA. EMBO J 18(21):6134–6145

Weng Y, Czaplinski K, Peltz SW (1996) Identification and characterization of mutations in the UPF1 gene that affect nonsense suppression and the formation of the Upf protein complex but not mRNA turnover. Mol Cell Biol 16(10):5491–5506

Wilschanski M, Yahav Y, Yaacov Y, Blau H, Bentur L, Rivlin J, Aviram M, Bdolah-Abram T, Bebok Z, Shushi L, Kerem B, Kerem E (2003) Gentamicin-induced correction of CFTR function in patients with cystic fibrosis and CFTR stop mutations. N Engl J Med 349(15):1433–1441

Wilson MA, Meaux S, van Hoof A (2007) A genomic screen in yeast reveals novel aspects of nonstop mRNA metabolism. Genetics 177(2):773–784

Wilusz CJ, Wilusz J (2005) Eukaryotic Lsm proteins: lessons from bacteria. Nat Struct Mol Biol 12(12):1031–1036

Wittmann J, Hol EM, Jack HM (2006) hUPF2 silencing identifies physiologic substrates of mammalian nonsense-mediated mRNA decay. Mol Cell Biol 26(4):1272–1287

Wurtmann EJ, Wolin SL (2009) RNA under attack: cellular handling of RNA damage. Crit Rev Biochem Mol Biol 44(1):34–49

Zhang S, Ruiz-Echevarria MJ, Quan Y, Peltz SW (1995) Identification and characterization of a sequence motif involved in nonsense-mediated mRNA decay. Mol Cell Biol 15(4):2231–2244

Zund D, Gruber AR, Zavolan M, Muhlemann O (2013) Translation-dependent displacement of UPF1 from coding sequences causes its enrichment in 3′ UTRs. Nat Struct Mol Biol 20(8):936–943

Chapter 9
Making Ribosomes: Pre-rRNA Transcription and Processing

Kathleen L. McCann and Susan J. Baserga

Abstract Ribosomes are essential, intricate cellular machines that translate mRNA into protein in all cells. The process of building a ribosome is highly complex. The ribosomal RNA (rRNA) has to be properly synthesized, processed, and assembled with ribosomal proteins in the cell nucleolus. Studies carried out in yeast have significantly increased our understanding of the mechanisms driving ribosome assembly in eukaryotes. This chapter will briefly review ribosome assembly in *Saccharomyces cerevisiae* and discuss what is known about ribosome assembly in other yeast species.

Contents

K. L. McCann · S. J. Baserga
Department of Genetics, Yale University School of Medicine, New Haven, CT 06520, USA

S. J. Baserga (✉)
Department of Molecular Biophysics and Biochemistry, Yale University
School of Medicine, New Haven, CT 06520, USA
e-mail: susan.baserga@yale.edu

S. J. Baserga
Department of Therapeutic Radiology, Yale University School of Medicine,
New Haven, CT 06520, USA

A. Sesma and T. von der Haar (eds.), *Fungal RNA Biology*,
DOI: 10.1007/978-3-319-05687-6_9,

Introduction

In all living organisms, ribosomes are the molecular machines that are responsible for translating mRNA into protein. As such, they are absolutely essential for cell viability. Eukaryotic ribosomes comprise two subunits, the 40S or small subunit (SSU) and the 60S or large subunit (LSU) (Fig. 9.1). The SSU contains the 18S ribosomal RNA (rRNA) and 33 ribosomal proteins (r-proteins) while the LSU contains the 5S, 5.8S, and 25S rRNAs and 46 r-proteins. Interestingly, it is the rRNA rather than the r-proteins that catalyzes peptide bond formation during protein synthesis (Chap. 5).

The assembly of ribosomes is highly complex and metabolically expensive. In rapidly growing *Saccharomyces cerevisiae*, over 2,000 ribosomes are assembled each minute (Warner 1999). To achieve this rapid rate of assembly, the cell utilizes all three RNA polymerases. RNA polymerase I and RNA polymerase III synthesize the pre-ribosomal RNAs (pre-rRNAs) while RNA polymerase II synthesizes the r-proteins and the numerous assembly proteins. Over 200 trans-acting assembly proteins are required for processing, folding, and assembly of the rRNAs with the r-proteins.

Due to the ease of genetic manipulation and the tractability of biochemistry, the model yeast *S. cerevisiae* has been paramount in examining ribosome assembly in eukaryotes. For example, many of the trans-acting factors required to make ribosomes were identified in yeast by tandem affinity purification and mass spectrometry (Bassler et al. 2001; Harnpicharnchai et al. 2001; Saveanu et al. 2001; Dragon et al. 2002; Fatica et al. 2002; Grandi et al. 2002). Epigenetic regulation of the rDNA was also first realized in yeast (reviewed in (Hamperl et al. 2013)). Furthermore, the first atomic model of a eukaryotic ribosome was solved in *S. cerevisiae* (Fig. 9.1) (Ben-Shem et al. 2011).

Ribosome assembly has been studied in other fungal species, albeit to a lesser extent. Pre-rRNA processing has been examined in *Schizosaccharomyces pombe* and *Candida albicans* (Good et al. 1997; Pendrak and Roberts 2011). Additionally, studies have been carried out in *S. pombe* to examine the sequence features of the pre-rRNA and their importance in pre-rRNA processing (reviewed in (Nazar 2004)). While many parts of the ribosome assembly pathway are conserved among yeast species and from yeast to humans, a few subtle species-specific differences have been recognized. As some yeast species are pathogenic and cause disease in humans (e.g. *C. albicans*), identification of differences in their ribosomes or ribosome assembly pathways is of medical importance. Currently, ribosomes are the primary target of antibiotics used to treat a variety of bacterial infections. These drugs selectively inhibit the activity of the bacterial ribosome (Lambert 2012). It is enticing to propose a similar logic for the development of novel antifungals. The remainder of this chapter will briefly summarize ribosome assembly in *S. cerevisiae* (for a more extensive review of ribosome biogenesis in *S. cerevisiae*, see (Woolford and Baserga 2013)), and discuss the work done to examine ribosome assembly in other yeast species.

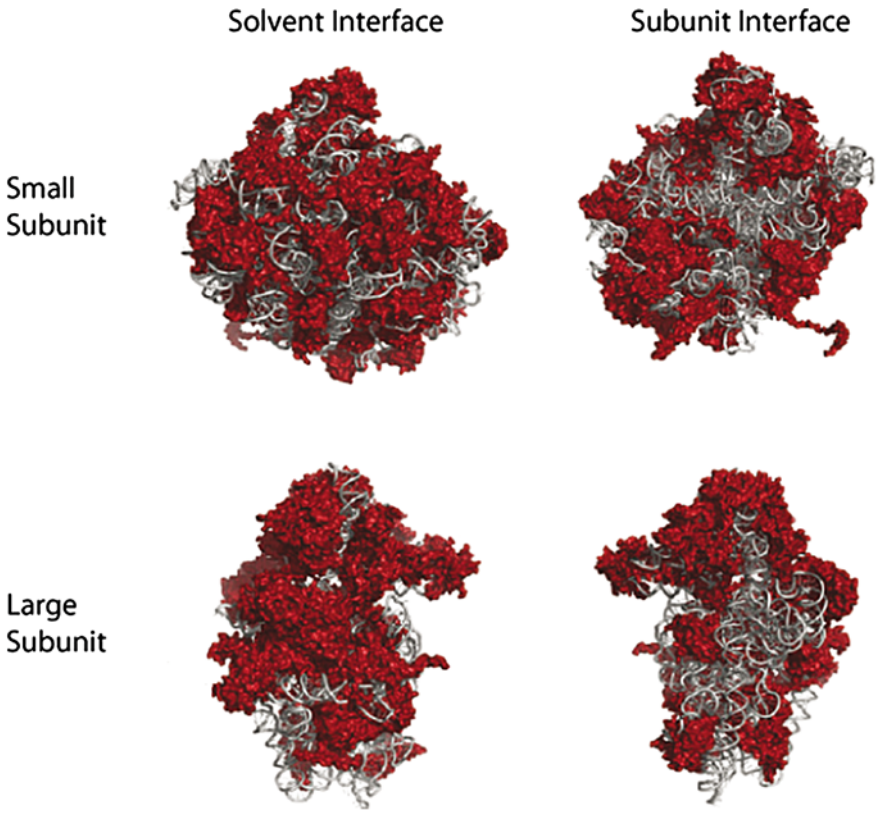


Fig. 9.1 *Crystal structure of the ribosomal subunits from S. cerevisiae at* 3.0 Å *resolution*. For both subunits, the solvent exposed surface is on the *left* while the subunit interface is on the *right*. The rRNA is shown in *gray* and the ribosomal proteins are *red*. The crystal structure is from (Ben-Shem et al. 2011). This figure was adapted from (Woolford and Baserga 2013)

Ribosome Assembly has been Extensively Described in *S. cerevisiae*

The process of making ribosomes occurs in the cell nucleolus, a non-membrane bound subcompartment of the nucleus. In eukaryotes, the nucleolus forms around the rDNA, which encodes the sequences for the mature 18S, 5.8S, and 25S rRNAs. Cells often carry multiple copies of the rDNA as transcription from one rDNA repeat is not sufficient to produce enough ribosomes to meet the demands of a growing cell. These copies, or rDNA repeats, are tandemly arrayed. In *S. cerevisiae*, roughly 150 rDNA repeats are arrayed on chromosome XII (Petes 1979).

Each rDNA repeat is transcribed by RNA polymerase I (RNA pol I) as a single unit, giving rise to a 35S polycistronic pre-rRNA precursor, which is roughly 6.6 kb in *S. cerevisiae* (Fig. 9.2). The 35S pre-rRNA contains the sequences for

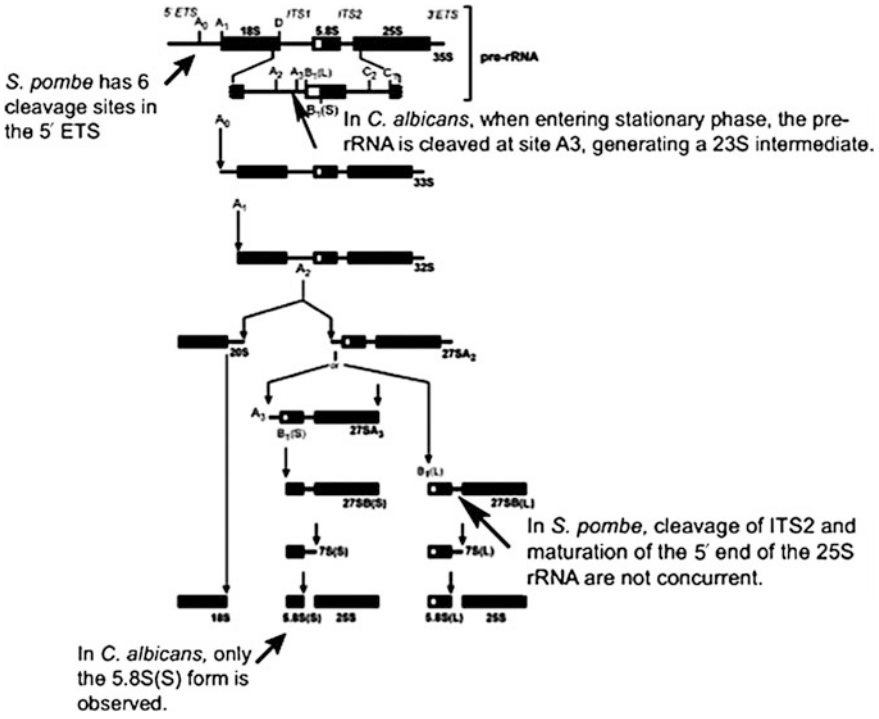


Fig. 9.2 *Pre-rRNA processing pathway in S. cerevisiae pre-rRNA*. The 35S pre-rRNA contains the 18S, 5.8S, and 25S rRNAs and several spacer sequences. The spacer sequences are removed by a series of endonucleolytic and exonucleolytic cleavage steps during the course of ribosome assembly. The major differences between pre-rRNA processing in *S. cerevisiae, S. pombe*, and *C. albicans* are highlighted. This figure was adapted from Granneman and Baserga (2004)

the 18S rRNA, which becomes part of the SSU, and the 5.8S and 25S rRNAs, which become part of the LSU, and several spacer sequences: the 5′ external transcribed spacer (5′ ETS), internal transcribed spacers 1 and 2 (ITS1 and ITS2) and the 3′ external transcribed spacer (3′ ETS). The 5S rRNA, which is a component of the LSU, is transcribed independently by RNA polymerase III in the opposite direction as the 35S pre-rRNA. During maturation of the rRNAs, the pre-rRNA is chemically modified and the spacer sequences are removed through a number of cleavage steps.

The pre-rRNAs are chemically modified in two ways: 2′-*O*-ribose methylation and pseudouridylation. Two classes of small nucleolar ribonucleoproteins (snoRNPs) are required to carry out the modifications that can occur cotranscriptionally (Osheim et al. 2004; Kos and Tollervey 2010). Catalysis of 2′-*O*-ribose methylation is performed by box C/D snoRNPs while box H/ACA snoRNPs catalyze pseudouridylation. Modification of the pre-rRNA by box C/D and box H/ACA snoRNPs is necessary for optimal function of mature ribosomes. The modified

nucleotides within the mature rRNAs are localized in regions that are important for ribosome function including the peptidyl transferase center, the polypeptide exit tunnel and the sites of subunit–subunit interaction (Decatur and Fournier 2002). Thus, ribosomes that are lacking these modifications exhibit reduced translational efficiency (Baxter-Roshek et al. 2007).

To generate the mature rRNAs, the external and internal transcribed spacers must be removed through a series of endonucleolytic and exonucleolytic processing steps (Fig. 9.2) (Venema and Tollervey 1995). In *S. cerevisiae*, processing of the 35S polycistronic pre-rRNA begins with an initial cleavage that is carried out by a large ribonucleoprotein, the SSU processome. Cleavage by the SSU processome can occur at sites A_0 or A_1, which lie in the 5′ ETS, or at site A_2, which lies in ITS1 (reviewed in (Phipps et al. 2011)). Cleavage at sites A_0 and A_1 remove the 5′ ETS and generate the 5′ end of what will become the mature 18S. After cleavage at site A_2, which separates the 20S pre-rRNA from the $27SA_2$ pre-rRNA, processing and assembly of the subunits proceeds independently in parallel pathways. Alternatively, in some instances, the 35S pre-rRNA is cleaved at site A_3 by RNase MRP. This cleavage generates the 23S and $27SA_3$ pre-rRNA intermediates (Lygerou et al. 1996). The 20S pre-rRNA is then exported to the cytoplasm where the remainder of ITS1 is removed by cleavage at site D to generate the mature 18S rRNA (Fatica et al. 2003).

Generation of the mature 5.8S and 25S rRNAs from the $27SA_2$ intermediate is more complex and occurs by two alternative pathways. Primarily, the $27SA_2$ pre-rRNA is cleaved by RNase MRP at site A_3 to generate the $27SA_3$ pre-rRNA intermediate. The 5′–3′ exonucleases Rat1 and Rrp17 digest the remaining ITS1 sequence until they reach site $B1_S$. This forms the 5′ end of the 5.8S within the $27SB_S$ pre-rRNA intermediate (Oeffinger et al. 2009). A small fraction of the time, cleavage of the $27SA_2$ site can occur at site $B1_L$, generating the $27SB1_L$ pre-rRNA intermediate. Processing of ITS2 from the $27SB_S$ and $27SB_L$ pre-rRNAs is indistinguishable. Cleavage in ITS2 occurs at site C2 to produce the 25.5S and $7S_S$ or $7S_L$ pre-rRNAs. To form the mature 25S, the 5′ end of the 25.5S pre-rRNA is trimmed by Rat1 (Geerlings et al. 2000). Processing of the $7S_S$ or the $7S_L$ to generate the mature 3′ end of the 5.8S is achieved in multiple steps by several 3′–5′ exonucleases including the exosome and Rrp6 (Briggs et al. 1998; Allmang et al. 1999; Faber et al. 2002).

The cleavage and processing steps required to make the mature rRNAs are carried out by a large number of assembly proteins (Tables 9.1 and 9.2). Ribosome assembly proteins were first identified through classic genetic approaches. However, the development of more powerful purification and mass spectrometric methods enabled the rapid biochemical identification of the majority of ribosome associated assembly factors (Bassler et al. 2001; Harnpicharnchai et al. 2001; Saveanu et al. 2001; Dragon et al. 2002; Fatica et al. 2002; Grandi et al. 2002; Woolford and Baserga 2013). Over 70 proteins have been identified as SSU assembly factors and over 90 proteins are thought to play a role in LSU assembly. The vast majority of these proteins are essential and conserved both among fungal species and to higher eukaryotes including humans (Tables 9.1 and 9.2). A number

Table 9.1 Assembly factors that function in SSU assembly are largely conserved among several yeast species

S. cerevisiae Gene Name	Essential in *S. cerevisiae*?	*S. pombe*	*C. albicans*	*N. crassa*	*A. nidulans*
Nop1/Lot3	Yes	fib1	NOP1	NCU03702/ fibrillarin	AN0745
Nop56/Sik1	Yes	SPBC646.10c	SIK1	NCU06943/ SIK1	AN10352
Nop58/Nop5	Yes	SPAC23G3.06	NOP5	NCU03396/ nop-58	AN3167/ nop58
Snu13	Yes	snu13	SNU13	NCU01331	AN1319
Rrp9	Yes	SPAC2E1P5.05	RRP9	NCU01367	AN4456
Imp3	Yes	imp3	IMP3	NCU01008/ IMP3	AN6171
Imp4	Yes	SPAC19A8.07c	IMP4	NCU04243/ IMP4	AN0647
Mpp10	Yes	SPAC23C11.03	MPP10	NCU03952/ Mpp10	AN4298
Utp4	Yes	utp4	UTP4	NCU00650/ Wdr1	AN3794
Utp5	Yes	utp5	UTP5	NCU08421	AN1983
Utp8	Yes		UTP8		
Utp9	Yes		UTP9		
Utp10	Yes	utp10	orf19.7215	NCU00336	AN5455
Utp15	Yes	utp15	UTP15	NCU09843	AN4648
Utp17/Nan1	Yes	SPAC1B3.13	NAN1	NCU02604	AN11162
Pol5	Yes	pol5	POL5	NCU02497	AN3944
Utp1/Pwp2	Yes	SPBC713.04c	PWP2	NCU03794	AN8183
Utp6	Yes	SPBC244.02c	orf19.2330	NCU05894	AN5010
Upt12/Dip2	Yes	dip2	DIP2	NCU03051	AN0305
Utp13	Yes	utp13	UTP13	NCU03628/ pod-5	AN4460
Utp28	Yes	SPBC29A3.06	UTP18	NCU03018	AN7458
Utp21	Yes	SPCC1672.07	UTP21	NCU07011	AN8293
Rrp7	Yes	SPBC776.17	orf19.5835	NCU04571	AN11929
Utp22	Yes	SPBC776.08c	UTP22	NCU09437	AN3455
Cka1	No	Cka1	CKA1	NCU03124/ cka	AN1485/ cka1
Cka2	No		CKA2		
Ckb1	No	ckb1	CKB1	NCU05485/ ckb-1	AN10546
Ckb2	No	ckb2	CKB2	NCU02754/ ckb-2	AN1272
Rrp36	Yes	SPAC823.04	orf19.2362	NCU03824	AN6561
Utp2/Nop14	Yes	nop14	NOP14	NCU07393	AN1434
Utp3/Sas10	Yes	SPBC3B8.09	SAS10	NCU00301/ Sas10/Utp3	AN1164
Utp7/Kre31	Yes	SPAC959.03c	orf19.4835	NCU01502	AN6265
Utp11	Yes	SPAC18G6.06	orf19.2386	NCU03133	AN0796
Utp14	Yes	SPAC57A7.06	orf19.7552	NCU07058	AN0475

(continued)

Table 9.1 (continued)

S. cerevisae Gene Name	Essential in *S. cerevisiae*?	*S. pombe*	*C. albicans*	*N. crassa*	*A. nidulans*
Utp16/Bud21	No	SPBP8B7.10c	BUD21		
Noc4/Utp19	Yes	SPBC1604.06c	NOC4	NCU00501	AN1533
Utp20	Yes	utp20	UTP20	NCU02494	AN3938
Utp23	Yes	utp23	orf19.3724	NCU02748	AN10052
Utp24/Fcf1	Yes	SPBC32H8.04c	orf19.3220	NCU04159	AN10865
Utp25	Yes	utp25	orf19.1849	NCU01022	AN1354
Utp30	No		orf19.154		
Bms1	Yes	bms1	BMS1	NCU04348	AN6334
Dbp8	Yes	dbp8	DBP8	NCU06418/ dbp-8	AN4903
Dhr1/Ecm16	Yes	SPAPB1A10.06c	orf19.2090	NCU03808/ PRP16	AN6585
Dhr2	Yes	prh1	orf19.107	NCU06318/ DHX8	AN4721
Emg1/Nep1	Yes	mra1	EMG1	NCU02428	AN2759
Krr1	Yes	mis3	KRR1	NCU07041/ mis3	AN5875
Rcl1	Yes	SPAC12G12.06c	RCL1	NCU02675	AN1474
Nop19/Dhi1	Yes		orf19.3970		
Rok1	Yes	rok1	CHR1	NCU00919/ rok-1	AN0944
Rrp3	Yes	SPAC823.08c	orf19.7546	NCU04504/ rrp-3	AN4233
Rrp5	Yes	SPCC1183.07	orf19.1578	NCU06272/ Rrp5	AN2147/ rrp5
Sof1	Yes	sof1	SOF1	NCU01595/ Sof1	AN4226
Dbp4/Hca4/ Ecm24	Yes	SPAC1093.05	HCA4	NCU04439/ dbp-4	AN0589
Enp1/Meg1	Yes	SPBC13G1.09	ENP1	NCU08323/ bystin	AN9398
Esf1	Yes	esf1	orf19.2319	NCU04485/ Esf1	AN7503
Esf2/Abt1	Yes	esf2	orf19.3161	NCU08668/ esf-2	AN7305
Fal1	Yes	SPAC1F5.10	FAL1	NCU01234	AN8016/ fal1
Fyv7	No	SPAC8C9.07	FYV5		
Gno1/Pxr1	Yes	SPAC890.05	orf19.3831	NCU02528	AN11066
Has1	Yes	has1	HAS1	NCU09349/ has-1	AN1949
Kre33	Yes	nat10	orf19.512	NCU02284/ Kre33	AN0247
Lcp5	Yes	SPAC18B11.06	orf19.7422	NCU07331/ Lcp5	AN5340
Ltv1/Ykl2	No	SPAC3F10.17	LTV1	NCU04165/ Ltv1	AN2262

(continued)

Table 9.1 (continued)

S. cerevisae Gene Name	Essential in *S. cerevisiae*?	*S. pombe*	*C. albicans*	*N. crassa*	*A. nidulans*
Mrd1	Yes	SPBP22H7.02c	orf19.1646	NCU02611	AN0421
Nop6	No	SPBC365.04c	NOP6	NCU02159	AN4748
Nop9	Yes	nop9	orf19.4479	NCU06118	AN10618
Nsr1/She5	No	gar2	orf19.6090	NCU03092	AN4865
Pfa1/Sqs1	No	SPBC30B4.02c	orf19.2400	NCU02434/ Sqs1	AN9061
Prp43/JA1	Yes	prp43	orf19.1687	NCU01612/ Prp43	AN0133
Sen1/Cik3/Nrd2	Yes	sen1	SEN1	NCU02036	AN8671
Sgd1	Yes	SPAC24C9.11	SGD1	NCU02406	AN4581
Slx9	No	Slx9	orf19.2237.1		

The following databases were used to generate the table for: *S. cerevisiae*: yeastgenome.org; *S. pombe*: pombase.org; *C. albicans*: candidagenome.org; *N. crassa*: broadinstitute.org/annotation/genome/neurospora/MultiHome.html; and *A. nidulans*: aspgd.org.

of assembly proteins have putative or known enzymatic activities including endonucleases, exonucleases, helicases, GTPases, AAA-ATPases, and kinases, but the precise function of the majority of assembly factors remains unknown. To this end, many assembly factors have been systematically depleted and assigned to a particular step in ribosome assembly based on their pre-rRNA processing defects (reviewed in (Woolford and Baserga 2013)). Furthermore, a subset of the assembly factors are known to form and function as discrete subcomplexes. For example, several subcomplexes are formed by known SSU biogenesis factors including the UtpA/t-Utp, UtpB, and UtpC subcomplexes (Dosil and Bustelo 2004; Gallagher et al. 2004; Krogan et al. 2004; Rudra et al. 2007).

Examination of Ribosome Assembly in Other Yeast Species

While ribosome assembly has been systematically examined in *S. cerevisiae*, only a disparate set of experiments have been done to describe ribosome biogenesis in other yeast species. The majority of these experiments have been performed in the other model yeast, *S. pombe,* although pre-rRNA processing in *C. albicans* has also been surveyed.

Investigation of ribosome assembly in *S. pombe* has offered novel mechanistic insight that is likely to be universal. For instance, analysis of the functional roles of pre-rRNA modification by 2′-*O*-ribose methylation and pseudouridylation revealed that chemical modifications significantly affect the stability of the mature rRNAs (Song and Nazar 2002). Furthermore, the role of the spacer sequences in pre-rRNA maturation has also been investigated in *S. pombe*. Computer modeling and nuclease protection assays have revealed important secondary structures within the

Table 9.2 Assembly factors that function in LSU assembly are largely conserved among several yeast species

S. cerevisae Gene Name	Essential in *S. cerevisiae*?	*S. pombe*	*C. albicans*	*N. crassa*	*A. nidulans*
Rrp5	Yes	SPCC1183.07	orf19.1578	NCU06272/ Rrp5	AN2147/ rrp5
Nop4/Nop77	Yes	SPBC4F6.14	NOP4	NCU06217	AN4330
Dbp6	Yes	SPCC285.03	orf19.3704	NCU11175	AN0637
Npa1/Urb1	Yes	SPCC14G10.02	orf19.2778	NCU16779	AN0237
Npa2/Urb2	Yes	urb2	orf19.5884	NCU06496	AN0659
Nop8	Yes	SPBP16F5.05c	NOP8		
Rsa3	No	rsa3	orf19.773		
Dbp2	Yes	dbp2	DBP2	NCU07839	AN5931
Noc1/Mak21	Yes	SPAC4F10.09c	MAK21	NCU09894	AN1200
Noc2/Rix3	Yes	SPAC1142.04	NOC2	NCU02066	AN0946
Noc3	Yes	noc3	orf19.7197	NCU00337	AN7260
Dbp3	No	dbp3	DBP3	NCU05782	AN7424
Dbp7	No	dbp7	DBP7	NCU06520	AN0204
Dbp9	Yes	SPCC1494.06c	orf19.3393	NCU07070	AN6374
Prp43/JA1	Yes	prp43	orf19.1687	NCU01612/ Prp43	AN0133
Loc1	No	SPAC1751.04	orf19.1642	NCU01301	AN6556
Mak5	Yes	SPBC4F6.07c	MAK5	ANU04041	AN1750
Cbf5	Yes	cbf5	orf19.962	NCU01290/ CBF5	AN8851/ swoC
Ssf1	No	SPAC1B9.03c	SSF1	NCU04503/ Ssf2	AN4232
Ssf2	No				
Rrp14	Yes	rrp14	orf19.2167	NCU06090	AN5702
Rrp15	Yes	SPAC227.02c	RRP15	NCU07795	AN8247
Mak16	Yes	mak16	MAK16	NCU04150	AN6612
Nop16	No	nop16	orf19.1388	NCU09528	AN7674
Ebp2	Yes	ebp2	orf19.4492	NCU01257	AN0074
Brx1	Yes	brx1	CSI2	NCU01524	AN10055
Rpf1	Yes	SPAC4F8.04	RPF2	NCU00404	AN8834
Ytm1	Yes	SPAC890.04c	YTM1	NCU03764	AN4469
Erb1	Yes	SPBC4F6.13c	ERB1	NCU03321	AN1367/ bop1
Nop7/Yph1	Yes	ppp1	PES1	NCU00925	AN3827
Drs1	Yes	ddx27	DRS1	NCU11175	AN10125
Nop15	Yes	SPCC1827.05c	NOP15	NCU08518	AN4346
Cic1/Nsa3	Yes	SPAC8F11.04	CIC1	NCU11321	AN6698
Rlp7	Yes	rlp7	RLP7	NCU07829	AN7107
Rrp1	Yes	nop52	orf19.6828	NCU07942	AN1175
Nop12	No	nop12	orf19.809	NCU09131	AN1452
Has1	Yes	has1	HAS1	NCU09349/ has-1	AN1949
Rai1	No	din1	orf19.6230	NCU04191	AN6514

(continued)

Table 9.2 (continued)

S. cerevisae Gene Name	Essential in *S. cerevisiae*?	*S. pombe*	*C. albicans*	*N. crassa*	*A. nidulans*
Rat1/Xrn2	Yes	dhp1	RAT1	NCU01643	AN0707
Xrn1	No	exo2	KEM1	NCU06678	AN11052
Rrp17	Yes	SPAC29A4.09	orf19.5066	NCU01203	AN5685
Nip7	Yes	SPCC320.11c	NIP7	NCU00105	AN8491
Nop2/Yna1	Yes	SPBP8B7.20c	orf19.501	NCU02511	AN8073
Spb4	Yes	SPBC24C6.02	SPB4	NCU03380	AN3176/ spb4
Rpf2	Yes	SPAC926.08c	RPF2	NCU08595	AN10200
Rrs1	Yes	rrs1	RRS1	NCU08121	AN3745
Dbp10	Yes	dbp10	orf19.5991	NCU02696	AN0583
Mak11	Yes	skb15	orf19.1791	NCU07864	AN4471
Tif6/Cdc95	Yes	tif6	orf19.1815	NCU09004	AN8824
Rlp24	Yes	rlp24	RLP24	NCU05235	AN4560
Nog1	Yes	nog1	NOG1	NCU05289	AN5865
Spb1	Yes	spb1	SPB1	NCU03669	AN0092
Nsa2	Yes	SPCP1E11.08	NSA2	NCU00981	AN2926/ nsa2
Nog2/Nug2	Yes	SPAC6F6.03c	NOG2	NCU02546	AN1666
Nug1	Yes	grn1	orf19.2917	NCU03066	AN1215
Cgr1	No	cgr1	orf19.2314	NCU02640	AN11145/ cgrA
Rix7	Yes	SPBC16E9.10c	RIX7	NCU01809	AN1366
Nsa1	Yes	SPBC83.15	NSA1		
Ipi3	Yes	crb3	orf19.6234	NCU00560	AN4771
Ipi2/Rix1	Yes	rix1	orf19.6862		AN2258
Ipi1	Yes	ipi1	orf19.475	NCU09094	AN8668
Rea1/Mdn1	Yes	SPCC737.08	MDN1	NCU06468	AN6310
Rsa4	Yes	SPCC18.05c	orf19.3778	NCU00794	AN10058
Mtr4/Dob1	Yes	mtr4	orf19.1335	NCU03363/ frh	AN4412/ mtr4
Nop53	Yes	rrp16	orf19.6886	NCU05233	AN10943
Sda1	Yes	sda1	SDA1	NCU09488	AN0844
Rrp12	Yes	rrp12	orf19.7011	NCU01675	AN6200
Arx1	No	arx1	ARX1	NCU00422	AN7299
Alb1	No	SPBC2D10.19c	orf19.7107	NCU01458	
Nmd3/Src5	Yes	nmd3	NMD3	NCU08663	AN1711
Mex67	Yes	mex67	MEX67	NCU09317	AN2737/ mexA
Mtr2	Yes	nxt1	MTR2		
Bud20	No		orf19.2934	NCU02364	AN5526
Ngl2	No	SPBC9B6.11c	orf19.5527		AN3602/ ccr4
Rex1/Rnh70	No	SPAC637.09	orf19.6195	NCU02961	AN7566
Rex2	No	rex2	REX2	NCU16377	AN10348

(continued)

Table 9.2 (continued)

S. cerevisae Gene Name	Essential in *S. cerevisiae*?	*S. pombe*	*C. albicans*	*N. crassa*	*A. nidulans*
Drg1/Afg2	Yes	SPBC56F2.07c	orf19.6432	NCU00018	AN7254
Rei1	No	SPCC550.15c	REI1		
Reh1	No	SPCC550.15c	REH1	NCU04022	AN1875
Jjj1	No	mug185	orf19.2399	NCU02432	AN9060
Ssa1	No	ssa1	SSA2	NCU09602/	AN5129/
Ssa2	No	ssa2		hsp70-1	hsp70
Yvh1	No	SPBC17A3.06	YVH1	NCU08158	AN4419
Mrt4	No	mrt4	MRT4	NCU07547	AN6902
Sdo1	Yes	sdo1	orf19.2708	NCU00476	AN5443
Efl1/Ria1	Yes	SPCC553.08c	RIA1	NCU06278	AN10856
Lsg1/Kre35	Yes	SPAC3F10.16c	orf19.3463	NCU04166	AN2263
Sqt1	Yes	SPAC25H1.08c	orf19.4029	NCU02477	AN7630
Puf6	No	puf6	orf19.3547	NCU09380	AN10173
Fpr4	No	fkbp39	Orf19.1030	NCU03241/	AN10489/
Fpr3	No			fkr-4	Fkbp5
Las1	Yes	las1	LAS1	NCU11261	AN6822
Grc3	Yes	grc3	orf19.721	NCU02237	AN4599
Rnt1	No	pac1	DCR1	NCU01762	AN11224

The following databases were used to generate the table for: *S. cerevisiae*: yeastgenome.org; *S. pombe*: pombase.org; *C. albicans*: candidagenome.org; *N. crassa*: broadinstitute.org/annotation/genome/neurospora/MultiHome.html; and *A. nidulans*: aspgd.org

spacer sequences that are critical for proper processing of these regions. The secondary structures observed in *S. pombe* appear to be conserved across eukaryotes despite the length and sequence differences in the spacer regions, thus reinforcing their significance (reviewed in (Nazar 2004)). Additionally, it has been suggested that the secondary structures bind trans-acting assembly proteins. Supporting this hypothesis, a large protein complex, ribosome assembly chaperone (RAC), has been copurified with both ITS1 and ITS2 sequences using affinity chromatography (Lalev et al. 2000; Lalev and Nazar 2001). RAC also binds sequences in the 3′ETS and promotes 3′ETS removal by Pac1, the ortholog of *S. cerevisiae* Rnt1, by inducing a conformational change (Spasov and Nazar 2008). Strikingly, when the sequences that bind RAC in ITS1 are mutated, processing of the pre-rRNA is significantly impaired. This suggests that cleavage in ITS1 is mediated by the RAC complex (Abeyrathne and Nazar 2005). Interestingly, 3′ETS removal has also been studied in *C. albicans* where it was found that Dicer is the functional equivalent of Rnt1 and is responsible for proper 3′ETS processing (Bernstein et al. 2012).

Analysis of pre-rRNA processing in *S. pombe* revealed that the major processing pathway is highly similar to that in *S. cerevisiae*. However, there are a few interesting, though subtle, differences. For example, processing of the 5′ETS is achieved through 6 cleavage steps in *S. pombe* compared to two cleavages in *S. cerevisiae*. Also, processing of ITS2 in *S. pombe* appears to occur through a

mechanism similar to processing of ITS1 whereby the spacer region is cleaved in the middle and then the 5′ end of the 25S is matured. This is different from ITS2 processing in *S. cerevisiae* where cleavage in ITS2 and 5′ end maturation of the 25S are concurrent (Fig. 9.2) (Raue and Planta 1995).

Pre-rRNA processing has also been examined in the pathogenic yeast, *C. albicans*. While the rDNA polycistronic structure of *C. albicans* is similar to that of other eukaryotes, the number of rDNA repeats can vary greatly depending on strain and growth conditions. As few as 21 or as many as 176 rDNA repeats have been identified on chromosome R (Rustchenko et al. 1993). The major pre-rRNA processing pathway in *C. albicans* is strikingly similar to that of *S. cerevisiae*. However, like *S. pombe*, there are a few distinguishing differences. In *C. albicans*, the 5′ end of the 5.8S is only found in one form that corresponds to the $5.8S_S$ in *S. cerevisiae* (Fig. 9.2) (Pendrak and Roberts 2011). Investigation of pre-rRNA processing in *C. albicans* also revealed the accumulation of a 23S precursor, suggestive of post-transcriptional processing, when cells entered stationary phase. Subsequent addition of nutrients resulted in the disappearance of the 23S intermediate and the resumption of canonical, co-transcriptional 18S maturation (Pendrak and Roberts 2011). This observation suggests a novel mechanism whereby *C. albicans* regulates rRNA processing in response to its growth conditions.

Interestingly, the 25S and 18S rRNAs were found to be polyadenylated in *C. albicans*. Polyadenylation of the rRNAs significantly increases when *C. albicans* shifts from yeast to its germinating forms (Fleischmann et al. 2004). Polyadenylation of the pre-rRNA is also observed in *S. cerevisiae* when the degradative exonuclease Rrp6p is depleted, thereby suggesting a role for polyadenylation in degradation of aberrant pre-rRNA precursors (Kuai et al. 2004). However, the role of polyadenylation in *C. albicans* has yet to be fully elucidated.

Proper processing of the pre-rRNA requires the function of over 200 trans-acting ribosome assembly factors, yet their role in ribosome assembly in species other than *S. cerevisiae* has been relatively understudied. Trans-acting ribosome biogenesis factors were first identified in *Aspergillus nidulans* in 1974 as mutants that exhibit cold sensitivity (Waldron and Roberts 1974). Cold sensitivity is common to ribosome assembly mutants as it is thought that their sub-optimal function renders them incapable of overcoming the higher energy barrier of ribosome synthesis imposed by colder temperatures (Guthrie et al. 1969). This phenomenon has been observed after depletion or mutation of ribosome assembly proteins in *S. cerevisiae* (Dragon et al. 2002; Gallagher and Baserga 2004; Freed and Baserga 2010). Despite the initial identification of several ribosome biogenesis proteins in *A. nidulans* almost 40 years ago, very little has been done since to examine their biochemical functions in ribosome assembly.

Many of the assembly proteins identified in *S. cerevisiae* appear to be conserved among yeast species (Tables 9.1 and 9.2). However, the functional assignment of these proteins as ribosome assembly factors is often based on sequence homology alone. This is especially true for the pathogenic yeasts *C. albicans* and *A. nidulans*. It is also striking that the majority of the trans-acting assembly proteins are

essential in *S. cerevisiae* (Tables 9.1 and 9.2). Their essentiality and conservation underscores the functional importance of these proteins in ribosome biogenesis.

However, not all essential proteins are conserved. For example, Utp8 is an essential member of the t-Utp subcomplex in *S. cerevisiae*, and it is not conserved in *S. pombe*, *N. crassa*, or *A. nidulans* (Tables 9.1 and 9.2). Interestingly, a Utp8 orthologue is also not found in humans, although a metazoan specific functional analog exists (Freed et al. 2012). The presence of a functional analog in humans suggests that similar analogs might exist in other yeast species. Thus, they may be good targets for the development of novel therapeutics to disrupt ribosome biogenesis in a species-specific manner.

Some biochemical analysis has been done on selected protein components in *S. pombe*. For example, the components of the RAC complex, which mediates cleavage of the pre-rRNA, have been identified (Lalev et al. 2000; Lalev and Nazar 2001). The function of Skb15 has also been examined in *S. pombe*. Mak11, the Skb15 homologue in *S. cerevisiae*, has a well-documented role in 60S assembly (Saveanu et al. 2007). Thus, the role of Skb15 in 60S biogenesis was investigated. Depletion of Skb15 led to a concomitant decrease in free 60S subunits. Also, the 7S and 27S precursors were enriched in Skb15 purifications compared to that of controls. Together, the data support a role for Skb15 in LSU assembly (Saveanu et al. 2007).

Perspectives

Ribosome biogenesis is a critical, intricate cellular process that requires the coordinated action of a large number of proteins and small RNAs. Much of our knowledge of this process in yeast comes from the extensive work done in *S. cerevisiae*, whose ease of growth, tractable genetics, and biochemistry make it a powerful model system. However, *S. pombe* has also been a useful model particularly in regard to revealing the importance of the spacer sequences in rRNA processing.

Given its importance for cell vitality and growth, the overall process of ribosome assembly is highly conserved in other eukaryotes; although, as examinations of pre-rRNA processing in *S. pombe* and *C. albicans* have revealed, there are subtle species-specific differences. It will be important to further delineate the steps in ribosome biogenesis in other yeast species, particularly those that are human pathogens. Any identified differences in their ribosome assembly pathways have the potential to become targets for the development of species-specific anti-fungals. Despite the high level of conservation, the identification of drugs that target bacterial ribosomes specifically have led to major advances in the treatment of bacterial infections. Thus, the making of ribosomes could be a target for novel treatments for pathogenic yeasts such as *C. albicans*, *A. nidulans*, and *C. neoformans*.

References

Abeyrathne PD, Nazar RN (2005) Parallels in rRNA processing: conserved features in the processing of the internal transcribed spacer 1 in the pre-rRNA from Schizosaccharomyces pombe. Biochemistry 44:16977–16987

Allmang C, Kufel J, Chanfreau G, Mitchell P, Petfalski E, Tollervey D (1999) Functions of the exosome in rRNA, snoRNA and snRNA synthesis. EMBO J 18:5399–5410

Bassler J, Grandi P, Gadal O, Lessmann T, Petfalski E, Tollervey D et al (2001) Identification of a 60S preribosomal particle that is closely linked to nuclear export. Mol Cell 8:517–529

Baxter-Roshek JL, Petrov AN, Dinman JD (2007) Optimization of ribosome structure and function by rRNA base modification. PLoS ONE 2:e174

Ben-Shem A, Garreau de Loubresse N, Melnikov S, Jenner L, Yusupova G, Yusupov M (2011) The structure of the eukaryotic ribosome at 3.0 A resolution. Science 334:1524–1529

Bernstein DA, Vyas VK, Weinberg DE, Drinnenberg IA, Bartel DP, Fink GR (2012) Candida albicans Dicer (CaDcr1) is required for efficient ribosomal and spliceosomal RNA maturation. Proc Natl Acad Sci USA 109:523–528

Briggs MW, Burkard KT, Butler JS (1998) Rrp6p, the yeast homologue of the human PM-Scl 100-kDa autoantigen, is essential for efficient 5.8 S rRNA 3′ end formation. J Biol Chem 273:13255–13263

Decatur WA, Fournier MJ (2002) rRNA modifications and ribosome function. Trends Biochem Sci 27:344–351

Dosil M, Bustelo XR (2004) Functional characterization of Pwp2, a WD family protein essential for the assembly of the 90 S pre-ribosomal particle. J Biol Chem 279:37385–37397

Dragon F, Gallagher JE, Compagnone-Post PA, Mitchell BM, Porwancher KA, Wehner KA et al (2002) A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. Nature 417:967–970

Faber AW, Van Dijk M, Raue HA, Vos JC (2002) Ngl2p is a Ccr4p-like RNA nuclease essential for the final step in 3′-end processing of 5.8S rRNA in Saccharomyces cerevisiae. RNA 8:1095–1101

Fatica A, Cronshaw AD, Dlakic M, Tollervey D (2002) Ssf1p prevents premature processing of an early pre-60S ribosomal particle. Mol Cell 9:341–351

Fatica A, Oeffinger M, Dlakic M, Tollervey D (2003) Nob1p is required for cleavage of the 3′ end of 18S rRNA. Mol Cell Biol 23:1798–1807

Fleischmann J, Liu H, Wu CP (2004) Polyadenylation of ribosomal RNA by Candida albicans also involves the small subunit. BMC Mol Biol 5:17

Freed EF, Baserga SJ (2010) The C-terminus of Utp4, mutated in childhood cirrhosis, is essential for ribosome biogenesis. Nucleic Acids Res 38:4798–4806

Freed EF, Prieto JL, McCann KL, McStay B, Baserga SJ (2012) NOL11, implicated in the pathogenesis of North American Indian childhood cirrhosis, is required for pre-rRNA transcription and processing. PLoS Genet 8:e1002892

Gallagher JE, Baserga SJ (2004) Two-hybrid Mpp10p interaction-defective Imp4 proteins are not interaction defective in vivo but do confer specific pre-rRNA processing defects in Saccharomyces cerevisiae. Nucleic Acids Res 32:1404–1413

Gallagher JE, Dunbar DA, Granneman S, Mitchell BM, Osheim Y, Beyer AL et al (2004) RNA polymerase I transcription and pre-rRNA processing are linked by specific SSU processome components. Genes Dev 18:2506–2517

Geerlings TH, Vos JC, Raue HA (2000) The final step in the formation of 25S rRNA in Saccharomyces cerevisiae is performed by 5′–>3′ exonucleases. RNA 6:1698–1703

Good L, Intine RV, Nazar RN (1997) The ribosomal-RNA-processing pathway in Schizosaccharomyces pombe. Eur J Biochem 247:314–321

Grandi P, Rybin V, Bassler J, Petfalski E, Strauss D, Marzioch M et al (2002) 90S pre-ribosomes include the 35S pre-rRNA, the U3 snoRNP, and 40S subunit processing factors but predominantly lack 60S synthesis factors. Mol Cell 10:105–115

Granneman S, Baserga SJ (2004) Ribosome biogenesis: of knobs and RNA processing. Exp Cell Res 296:43–50

Guthrie C, Nashimoto H, Nomura M (1969) Structure and function of E. coli ribosomes. 8. Cold-sensitive mutants defective in ribosome assembly. Proc Natl Acad Sci USA 63:384–391

Hamperl S, Wittner M, Babl V, Perez-Fernandez J, Tschochner H, Griesenbeck J (2013) Chromatin states at ribosomal DNA loci. Biochim Biophys Acta 1829:405–417

Harnpicharnchai P, Jakovljevic J, Horsey E, Miles T, Roman J, Rout M et al (2001) Composition and functional characterization of yeast 66S ribosome assembly intermediates. Mol Cell 8:505–515

Kos M, Tollervey D (2010) Yeast pre-rRNA processing and modification occur cotranscriptionally. Mol Cell 37:809–820

Krogan NJ, Peng WT, Cagney G, Robinson MD, Haw R, Zhong G et al (2004) High-definition macromolecular composition of yeast RNA-processing complexes. Mol Cell 13:225–239

Kuai L, Fang F, Butler JS, Sherman F (2004) Polyadenylation of rRNA in Saccharomyces cerevisiae. Proc Natl Acad Sci USA 101:8581–8586

Lalev AI, Abeyrathne PD, Nazar RN (2000) Ribosomal RNA maturation in Schizosaccharomyces pombe is dependent on a large ribonucleoprotein complex of the internal transcribed spacer 1. J Mol Biol 302:65–77

Lalev AI, Nazar RN (2001) A chaperone for ribosome maturation. J Biol Chem 276:16655–16659

Lambert T (2012) Antibiotics that affect the ribosome. Rev Sci Tech 31:57–64

Lygerou Z, Allmang C, Tollervey D, Seraphin B (1996) Accurate processing of a eukaryotic precursor ribosomal RNA by ribonuclease MRP in vitro. Science 272:268–270

Nazar RN (2004) Ribosomal RNA processing and ribosome biogenesis in eukaryotes. IUBMB Life 56:457–465

Oeffinger M, Zenklusen D, Ferguson A, Wei KE, El Hage A, Tollervey D et al (2009) Rrp17p is a eukaryotic exonuclease required for 5′ end processing of Pre-60S ribosomal RNA. Mol Cell 36:768–781

Osheim YN, French SL, Keck KM, Champion EA, Spasov K, Dragon F et al (2004) Pre-18S ribosomal RNA is structurally compacted into the SSU processome prior to being cleaved from nascent transcripts in Saccharomyces cerevisiae. Mol Cell 16:943–954

Pendrak ML, Roberts DD (2011) Ribosomal RNA processing in Candida albicans. RNA 17:2235–2248

Petes TD (1979) Yeast ribosomal DNA genes are located on chromosome XII. Proc Natl Acad Sci USA 76:410–414

Phipps KR, Charette J, Baserga SJ (2011) The small subunit processome in ribosome biogenesis-progress and prospects. Wiley Interdiscip Rev RNA 2:1–21

Raue HA, Planta RJ (1995) The pathway to maturity: processing of ribosomal RNA in Saccharomyces cerevisiae. Gene Expr 5:71–77

Rudra D, Mallick J, Zhao Y, Warner JR (2007) Potential interface between ribosomal protein production and pre-rRNA processing. Mol Cell Biol 27:4815–4824

Rustchenko EP, Curran TM, Sherman F (1993) Variations in the number of ribosomal DNA units in morphological mutants and normal strains of Candida albicans and in normal strains of Saccharomyces cerevisiae. J Bacteriol 175:7189–7199

Saveanu C, Bienvenu D, Namane A, Gleizes PE, Gas N, Jacquier A et al (2001) Nog2p, a putative GTPase associated with pre-60S subunits and required for late 60S maturation steps. EMBO J 20:6475–6484

Saveanu C, Rousselle JC, Lenormand P, Namane A, Jacquier A, Fromont-Racine M (2007) The p21-activated protein kinase inhibitor Skb15 and its budding yeast homologue are 60S ribosome assembly factors. Mol Cell Biol 27:2897–2909

Song X, Nazar RN (2002) Modification of rRNA as a 'quality control mechanism' in ribosome biogenesis. FEBS Lett 523:182–186

Spasov K, Nazar RN (2008) RAC protein induces enzymatic access to the maturing 3′-end of the 25S rRNA in Schizosaccharomyces pombe. Biochem Biophys Res Commun 368:374–378

Venema J, Tollervey D (1995) Processing of pre-ribosomal RNA in Saccharomyces cerevisiae. Yeast 11:1629–1650

Waldron C, Roberts CF (1974) Cold-sensitive mutants in Aspergillus nidulans. II. Mutations affecting ribosome production. Mol Gen Genet 134:115–132

Warner JR (1999) The economics of ribosome biosynthesis in yeast. Trends Biochem Sci 24:437–440

Woolford JL Jr, Baserga SJ (2013) Ribosome Biogenesis in the Yeast Saccharomyces cerevisiae. Genetics 195:643–681

Chapter 10
Biogenesis and Evolution of Functional tRNAs

David Piñeyro, Adrian Gabriel Torres and Lluis Ribas de Pouplana

Abstract Faithful translation of genomes into proteomes depends, mainly, on the activity of transfer RNA (tRNA) as universal adaptor, as insightfully predicted by Crick. The central role of this relatively simple oligonucleotide depends upon a very large number of intermolecular interactions, some of which require that tRNAs maintain a constant general structure, while others depend on specific features that discriminate any given tRNA from the rest. Posttranscriptional modifications that increase the chemical diversity contained in the nucleotides of tRNAs can serve both purposes. Chemical modifications of tRNAs, thus, come in two general flavors: those that help to maintain the general shape of the molecule, and those that improve its interactions with one or more of its many molecular partners. Although the function of most of the chemical modifications known to occur in tRNAs remain unknown, up-to-date knowledge allows us to analyze the majority of them in some model organisms, including *Saccharomyces cerevisiae*. Here we will review our current understanding on the function of tRNA modifications, generally dividing them into two families: those that are likely to influence the structure of tRNA, and those that may play a role in the codon-anticodon interaction at the decoding center of the ribosome.

Contents

D. Piñeyro · A. G. Torres · L. R. de Pouplana (✉)
Institute for Research in Biomedicine (IRB Barcelona), Barcelona, Catalunya, Spain
e-mail: lluis.ribas@irbbarcelona.org

L. R. de Pouplana
Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Catalunya, Spain

A. Sesma and T. von der Haar (eds.), *Fungal RNA Biology*,
DOI: 10.1007/978-3-319-05687-6_10,

Introduction

Faithful translation of genomes into proteomes depends, mainly, on the activity of transfer RNA (tRNA) as universal adaptor, as insightfully predicted by Crick (Crick 1963). The central role of this relatively simple oligonucleotide depends upon a very large number of intermolecular interactions, some of which require that tRNAs maintain a constant general structure, while others depend on specific features that discriminate any given tRNA from the rest. Posttranscriptional modifications that increase the chemical diversity contained in the nucleotides of tRNAs can serve both purposes.

Chemical modifications of tRNAs, thus, come in two general flavors: those that help to maintain the general shape of the molecule, and those that improve its interactions with one or more of its many molecular partners. Although the function of most of the chemical modifications known to occur in tRNAs remain unknown, up-to-date knowledge allows us to analyze the majority of them in some model organisms, including *Saccharomyces cerevisiae*. Here we will review our current understanding on the function of tRNA modifications, generally dividing them into two families: those that are likely to influence the structure of tRNA, and those that may play a role in the codon-anticodon interaction at the decoding center of the ribosome.

The structural role played by chemical modifications in tRNA was clearly demonstrated by Kowalak and colleagues (Kowalak et al. 1994), who showed that, in the hyperthermophile *Pyrococcus furiosus*, the levels of tRNA modifications progressively increased with culture temperature. Most of the modifications that exhibited an apparent role in the adaptation to extreme temperatures are, in fact, unique to the archaeal hyperthermophiles. Similarly, in yeast, modifications are used to improve tRNA stability. For example, mutations that affect the activity of Trm6/Trm61, responsible for the formation of 1-methyl adenosine, lead to a reduction in steady-state levels of mature $tRNA^{iMet}$ due to a decrease in the stability of the molecule (Anderson et al. 1998).

More challenging remains the analysis of the function of modifications in the anticodon, given the complexity of the interactions that depend on the structure of this region of tRNA. In this regard, the analysis by mass spectrometry in yeast of general variations in tRNA modifications upon stress induction has revealed a complex and fluid pattern of adaptation that involves several modifications and positions in the tRNA (Chan et al. 2010, 2012). These ground-breaking experiments have demonstrated that the levels of certain modification enzymes are regulated as a response to specific insults, possibly because their activity is important in the control of the cellular responses to that particular stress. In the coming years we expect that these regulatory mechanisms will become clear, not only from the point of view of the regulation of the expression of these enzymes, but also in terms of the specific roles that each of the involved modifications play in the shaping of the genetic program and the proteome.

The connections described between tRNA modifications and human disease continue to grow (Torres et al. 2014) and, with them, the interest in the functional role played by these chemical groups in cellular homeostasis. Particularly, exciting is the possibility that anticodon modifications may play a role in the regulation of translation of certain genes. We have recently shown that highly expressed genes in yeast are enriched in codons whose translation requires tRNAs with inosine at position 34 (I34) in the anticodon (Novoa et al. 2012). I34 is generated after deamination of A34 by the enzyme adenosine deaminase acting on tRNA (ADAT), which is one of the few modifications known to be essential in yeast (Gerber and Keller 1999). The open question now is deciphering which genetic programs might be controlled by variations in levels of tRNA modifications, and what are the dynamics of the same modifications in tRNA populations.

tRNA Biogenesis

tRNA biogenesis is a complex process that leads to the production of mature and functional tRNA molecules. There are many proteins involved in this process which spans along several steps and cellular compartments. The entire biosynthetic pathway comprises the following key steps: transcription of the pre-tRNA molecule from its gene, removal of the 5′ leader and 3′ trailer sequences, addition of a 3′-terminal CCA motif, splicing of introns that may be present and chemical modification of particular nucleoside residues (Fig. 10.1).

In eukaryotes, tRNAs can be nuclear- or organelle-encoded (i.e., encoded in genome containing organelles such as mitochondria and chloroplasts). This section is focused on cytosolic tRNAs produced from nuclear genes. We will mainly focus on the process of tRNA biogenesis in *S. cerevisiae*, but the most relevant differences with other organisms will be mentioned where appropriate.

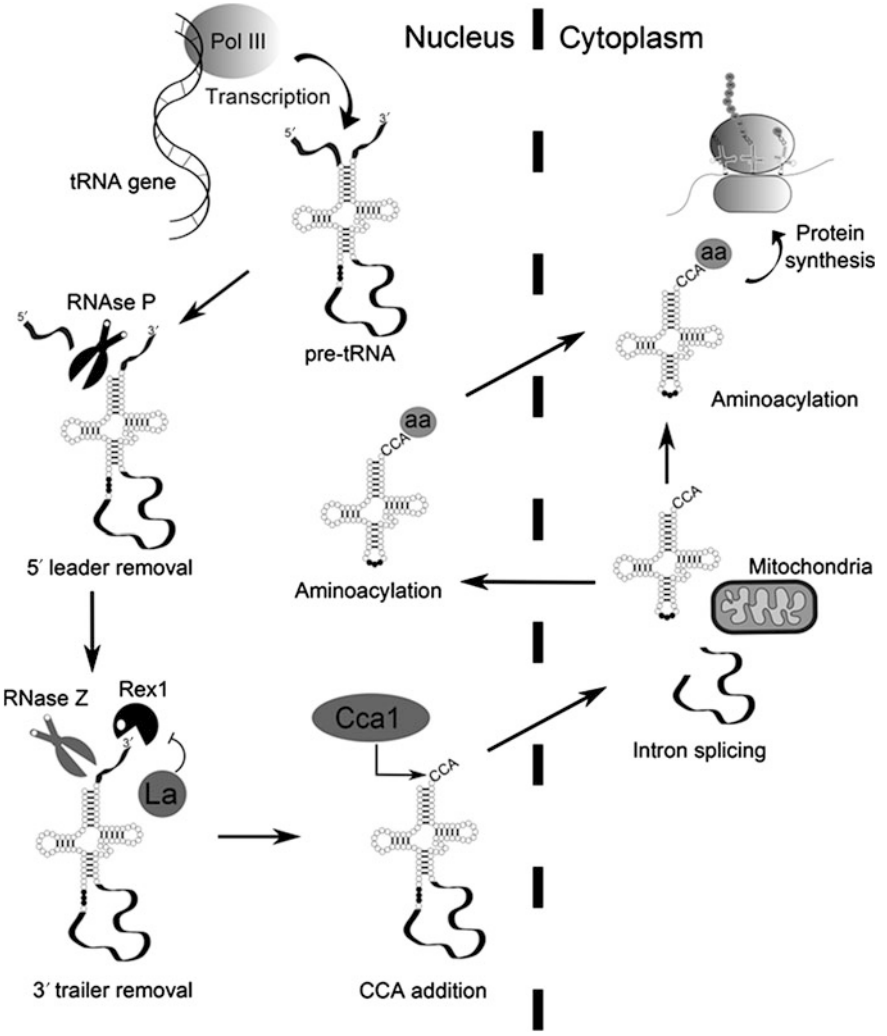


Fig. 10.1 tRNA biogenesis. tRNA biogenesis begins in the nucleolus, with the transcription of the pre-tRNA molecule. Afterward, the 5′ leader sequence (*ribbon*) is removed by the RNase P ribonucleoprotein (*black scissors*). Then, RNase Z (*gray scissors*), Rex1 (*Pacman symbol*), and La protein (*gray circle*) perform the 3′ trailer (*ribbon*) removal. The binding of La protein prevents the exonucleolytic activity of Rex1 and leads to endonucleolytic scission by RNase Z. Once the 5′ and 3′ tails have been removed, the CCA nucleotidyl transferase (Cca1, *gray ellipse*) adds the CCA tail. After that, the pre-tRNA molecule is transported to the cytoplasm. Intron splicing takes place at the mitochondrial outer membrane surface. The now mature tRNA molecule is aminoacylated by its cognate aminoacyl-tRNA synthetase (ARS) in the cytoplasm and incorporated to the protein synthesis. Alternatively, under certain conditions, the mature tRNA can be imported back into the nucleus, where it can be aminoacylated by a nuclear pool of ARSs and then re-exported to the cytoplasm. The pre-tRNA residues undergo several chemical modifications throughout the maturation process

tRNA Transcription

Nuclear tRNA genes are transcribed by RNA polymerase III (RNA pol III) and its associated machinery. In terms of promoter complexity and diversity, RNA pol III stands at an intermediate position compared to the rest of the RNA polymerases, since it is responsible for the transcription of a limited number of different RNA molecules: all the nuclear encoded tRNAs, the 5S ribosomal RNA (rRNA), U6 small nuclear RNA (snRNA), and many other nonprotein-coding RNAs (ncRNAs) (Willis 1993; Dieci et al. 2007; White 2011). Transcription by RNA pol III depends on several cis-acting elements (i.e., sequence elements) and some transcription factors (TFs). In the case of tRNA genes, the required cis-acting elements are two internal regions of the genes known as the A and B boxes. The A box is usually located about 10–20 bp downstream of the Transcription Start Site (TSS), and the B box normally sits at 30–60 bp downstream of the A box (Willis 1993; Dieci et al. 2007; Dumay-Odelot et al. 2010; Orioli et al. 2012). These sequences are involved in the correct positioning of the necessary transcription factor III C (TFIIIC), and the formation of the conserved D-loop and T-loop of the tRNA structure (Orioli et al. 2012). Additionally, flanking sequences upstream of the TSS are also important in some organisms, a TATA box being the most prevalent one (Geiduschek and Kassavetis 2001; Hamada et al. 2001). Although absent in *S. cerevisiae*, many tRNA promoters of *Schizosaccharomyces pombe* contain a TATA box at $\sim$30 bp upstream of the TSS, which is essential for the expression of the tRNA gene (Hamada et al. 2001).

The transcription process begins with the binding of the multi-subunit TFIIIC to the A and B boxes of the tRNA gene. The B box is known to be the most important promoter element for such interaction as TFIIIC has been found bound to isolated B box elements, but not to isolated A box elements (Noma et al. 2006; Wallrath and Geyer 2006; Moqtaderi et al. 2010). However, isolated B boxes are usually nontranscriptionally active, with only few exceptions producing ncRNAs of yet unknown function (Olivas et al. 1997; Parrott and Mathews 2007; Parrott et al. 2011). Once TFIIIC is on the tRNA gene, it recruits a second essential transcription factor, TFIIIB, which is a multi-protein complex composed of three different subunits: TATA box Binding Protein (TBP), Brf1, and Bdp1 (Kassavetis et al. 1991; Willis 2002). The finding that TBP is a universal component for all of the nuclear transcription machineries highlighted their common origin and conservation (Hernandez 1993; Rigby 1993; White et al. 1992). Moreover, the Brf1 amino terminal part is also related with TFIIB, a component of the RNA pol II machinery (Kassavetis et al. 1998). This conservation also extends to the polymerase itself, as five subunits are shared by the three main eukaryotic RNA polymerases (Carles et al. 1991; Vannini and Cramer 2012). Once TFIIIB is bound upstream of the TSS, it participates in the promoter opening and the recruitment of RNA pol III to start transcription (Acker et al. 2013). The transcription termination signal is relatively simple, consisting of a stretch of thymidine (T) residues in the coding sequence. The length of the stretch varies between species, being $\geq$4T in

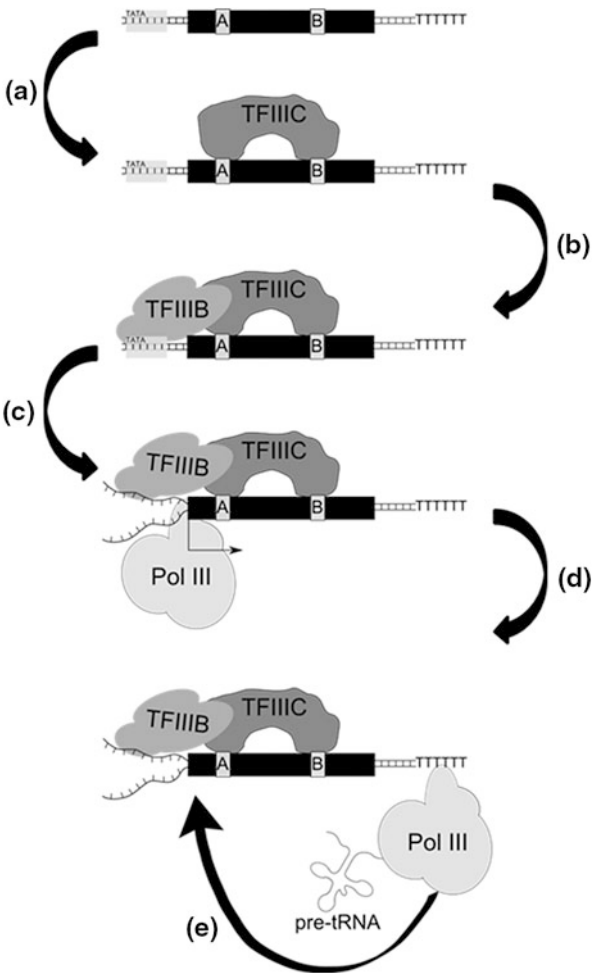


Fig. 10.2 tRNA transcription**. a** tRNA transcription begins with the binding of the TFIIIC complex to the A and B boxes in the tRNA gene sequence. **b** TFIIIC recruits the TFIIIB complex. TFIIIB binds upstream the transcription start site (TSS), which may contain a TATA box. **c** TFIIIB participates in the promoter opening and the recruitment of the RNA pol III (Pol III) complex, to initiate the transcription. **d** The transcription ends when Pol III finds the termination signal, consisting of a stretch of thymidines (T). **e** Finally, Pol III can undergo several rounds of reinitiation

vertebrates, ≥5T in *S. pombe*, and ≥6T in *S. cerevisiae*. It is worth noting that the efficiency of the termination signal can be strongly influenced by the surrounding sequences (Arimbasseri et al. 2013). That kind of weak terminators, when found between two transcription units, are thought to bring new opportunities for transcription regulation (Orioli et al. 2012). Finally, RNA pol III can undergo several rounds of reinitiation which, at least in fast growing cells, ensures the necessary yield of tRNAs (Dieci et al. 2013) (Fig. 10.2).

The study of the regulation of tRNA transcription is a relatively new field. The Maf1 protein was the first described and, to date, the only global negative regulator of RNA pol III transcription (Boguta et al. 1997). The activity of Maf1 is dependent on its phosphorylation state, thus connecting diverse unfavorable growth conditions with RNA pol III activity (Oficjalska-Pham et al. 2006; Roberts

et al. 2006). Regarding the RNA pol III activators, the Sub1 protein has a strong positive effect on RNA pol III transcription in vitro (Wang and Roeder 1998). Sub1 is the yeast homologue of the mammalian co-activator factor PC4, involved in the activation of many RNA pol II genes (Werten and Moras 2006; Malik et al. 1998; Ge and Roeder 1994). It has been shown that Sub1 participates both in the initiation and reinitiation of RNA pol III transcription (Tavenet et al. 2009). In addition to these two factors, some of the components of the RNA pol III transcription apparatus are phosphoproteins, and some kinases involved in their phosphorylation have been identified (Acker et al. 2013). Moreover, recent genome-wide studies established a link between chromatin state and RNA pol III activity (Brogaard et al. 2012; Mavrich et al. 2008). In this regard, it has been shown that several chromatin remodeling complexes such as FACT, Nhp6, and Isw2 are involved in RNA pol III transcription regulation, affecting in many cases the transcription of tRNA genes (see also Chap. 13). Likewise, the histone acetylation state also affects the expression of RNA pol III genes, thus linking histone acetylases and deacetylases to this process (Acker et al. 2013). Finally, many other proteins, most of them being known RNA pol II TFs, have been identified associated with RNA pol III genes by chromatin immunoprecipitation (ChIP) and subsequent genome-wide studies. However, their actual implication in the regulation of those RNA pol III genes remains unclear (Acker et al. 2013).

Trimming of the 5′ Leader and 3′ Trailer Ends

For most eukaryotic tRNA genes transcription starts some bases before and ends some bases after the limits of the mature tRNA, resulting in a pre-tRNA molecule with 5′ leader and 3′ trailer sequences. In *S. cerevisiae* the length of each of these extra sequences is about 12 bases, and its trimming is a stepwise process where the 5′ leader excision usually precedes the 3′ trailer removal (O'Connor and Peebles 1991; Hopper 2013) (Fig. 10.1).

The 5′ end processing is carried out by the ribonucleoprotein RNase P. This ancient ribozyme is present in all the organisms and organelles and is thought to be a relic from the "RNA world", the proposed evolutionary age when the enzymatic reactions were conducted by nucleic acids instead of proteins (Ellis and Brown 2009). The RNase P enzymatic activity is carried out by its RNA component, conserved in all life domains. There are a few rare cases where RNase P RNA has been substituted by proteins, as in human mitochondrial RNase P or the nuclear and mitochondrial RNase P of higher plants (Holzmann et al. 2008; Walker and Engelke 2008; Thomas et al. 2000; Gutmann et al. 2012). Besides its contribution to tRNA processing, RNase P is also implicated in the maturation of other RNA precursors such as small nucleolar RNAs (snoRNAs) and the *HRA1* ncRNA in yeast (Coughlin et al. 2008; Samanta et al. 2006; Yang and Altman 2007).

The 3′ end processing is more complex and involves both endo- and exonuclease activities (Phizicky and Hopper 2010). In yeast, three proteins seem to drive

this process: Rex1, RNase Z, and La protein (Hopper 2013). The 3′-5′ exonuclease activity is performed by Rex1, also involved in the processing of other ncRNAs (e.g., 5S rRNA, 5.8S rRNA, and snRNAs (van Hoof et al. 2000; Copela et al. 2008; Ozanick et al. 2009). The RNase Z (Trz1 in yeast) protein is in charge of the endonuclease activity, and processes both nuclear and mitochondrial pre-tRNAs (Chen et al. 2005; Daoud et al. 2012; Maraia and Lamichhane 2011). It has been proposed that the two aforementioned nucleases have differential access to the tRNA depending on the previous binding of the La protein (Lhp1), which prevents Rex1 activity and leads to endonucleolytic maturation by RNAse Z (Yoo and Wolin 1997).

Additions to Pre-tRNA Ends: CCA and tRNA^{His} G$_{-1}$

In eukaryotes, after the 3′ trailer removal, the required 3′ CCA sequence is added (Fig. 10.1). This process is accomplished by the CCA nucleotidyl transferase, Cca1 in yeast (Aebi et al. 1990), which is well conserved in different organisms. It is noteworthy that, in *Escherichia coli*, tRNA genes have the 3′ CCA encoded in the genomic sequence and therefore do not need the nucleotidyl transferase activity. However, they do conserve a homologous gene that, when deleted, leads to a slow growth rate phenotype, suggesting a function in tRNA 3′ end repair (Zhu and Deutscher 1987). The *CCA1* gene in yeast encodes for several splicing isoforms, each of them with distinct subcellular localization. It has been shown that the CCA adding activity is performed by the nuclear isoform (Wolfe et al. 1996), whereas the cytosolic isoform seems to be involved in tRNA end repair (potentially being the functional yeast equivalent to the *E. coli* enzyme) (Wolfe et al. 1994).

Another terminal nucleotide addition is necessary for the maturation of eukaryotic tRNA^{His}. After 5′ end trimming, this tRNA undergoes the addition of a G to its 5′ end (position −1 or G$_{-1}$) in order to be efficiently aminoacylated by histidine-tRNA synthetase (HisRS) (Rudinger et al. 1994). The enzyme responsible for this activity is Thg1, which is an essential gene with 3′-5′ nucleotide addition capacity, contrary to the regular nucleotide polymerization direction (Gu et al. 2003, 2005). While Thg1 null mutants are lethal, overexpression of both tRNA^{His} and HisRS under this genetic background can give a viable but unfit yeast strain (Preston and Phizicky 2010).

Intron Splicing

All eukaryotes and archaea sequenced to date carry introns in some of their tRNA genes, and their removal seems to be essential to accomplish the complete set of tRNAs. The length of introns is variable and, in fungi, ranges from just over 10 to about 60 nucleotides (Chan and Lowe 2009). The intron presence is not well

conserved throughout eukaryotes, with the exception of the $tRNA^{Tyr}$, which almost always contains introns. Perhaps the reason for this particular conservation lies in the fact that the intron processing is essential for subsequent modifications, as it was found in yeast (Johnson and Abelson 1983) (discussed in section "Chemical modifications on tRNAs"). The introns of all eukaryotic tRNA genes are always located between positions 37 and 38, whereas in archaea they can sometimes be located at other positions (Chan and Lowe 2009).

The splicing process of tRNA is simpler than mRNA splicing as carried out by the spliceosome (revised in Chap. 2), as there are less proteins involved. In yeast, the overall process consists of three steps, starting with the cleavage of the intron by the tRNA splicing endonuclease, resulting in two tRNA half molecules (Trotta et al. 1997, 2006). The second step is accomplished by the tRNA ligase (Trl1) and consists on the ligation of the two tRNA exons (Phizicky et al. 1986). This is a complex reaction that leaves a residual 2′ phosphate at the splice junction that has to be removed by the last enzyme, the phosphotransferase Tpt1 (Spinelli et al. 1997).

One of the most striking and unexpected features of the yeast tRNA splicing machinery is its subcellular localization. Whereas in vertebrates this process is accomplished in the nucleus (Melton et al. 1980), in yeast it takes place in the cytoplasm (Yoshihisa et al. 2007) (Fig. 10.1). The localization of the three enzymes required for the splicing process described above gave a clue in this regard. In particular, the tRNA endonuclease is located on the outer part of the mitochondria membrane (Yoshihisa et al. 2003; Huh et al. 2003), whereas the tRNA ligase is located in the cytoplasm (Huh et al. 2003; Mori et al. 2010), and the phosphotransferase is distributed between nucleus and cytoplasm (Dhungel and Hopper 2012). Strong evidence indicates that the tRNA splicing occurs on the mitochondrial surface, as mutants of the tRNA endonuclease enzyme accumulated unspliced pre-tRNAs in the cytoplasm (Yoshihisa et al. 2007), as well as mutants of the tRNA nuclear export machinery were deficient in tRNA splicing (Hopper and Banks 1978; Hopper et al. 1980). Interestingly, experiments using yeast engineered to have the tRNA splicing machinery only in the nucleus reveal that, although they processed the tRNA correctly, they cannot grow without the mitochondrial component of the tRNA endonuclease. This result suggests that the peculiar localization of the tRNA splicing machinery in yeast could be explained by other functions in which this machinery may be involved (Dhungel and Hopper 2012).

Modifications During tRNA Maturation

Posttranscriptional modifications on the tRNA (that will be discussed in depth in section "Chemical modifications on tRNAs") occur during its maturation process. Experiments carried out in *Xenopus* oocytes showed that the intricate series of modifications that pre-tRNA undergoes are introduced in an ordered manner (Nishikura and De Robertis 1981). This seems also to be true in yeast. For

example, in $tRNA^{Phe}_{GAA}$, some modifications can only occur in the presence of the tRNA intron (e.g., 5-methylcytidine at position 40); while others require a spliced tRNA (e.g., 2′O-methylcytidine, 2′O-methylguanosine, and 1-methylguanosine at positions 32, 34, and 37, respectively) (Jiang et al. 1997). However, some biochemical approaches suggest that the order of the tRNA modifications observed in vivo cannot be entirely explained by substrate specificity since it is usually possible to modify an in vitro produced tRNA molecule with the appropriate in vitro purified modification enzyme (Hopper and Phizicky 2003). Therefore, the subcellular localization of the tRNA modification enzymes seems to be an important factor in regulating the stepwise process of tRNA modification. In yeast, the subcellular localization of some tRNA modification enzymes is known. For example, N2, N2-dimethylguanosine at position 26 is produced in the nucleus by Trm1, and 2-methylguanosine at position 10 is made by Trm11 in the cytoplasm (Purushothaman et al. 2005; Rose et al. 1995).

Quality Control and tRNA Decay

tRNAs undergo several quality control mechanisms which ensure their correct function, as well as their regulated turnover during certain stress conditions. Two independent tRNA degradation pathways have been described: the nuclear exosome pathway and the rapid tRNA decay (RTD) pathway (Hopper 2013; Phizicky and Hopper 2010; Parker 2012).

The exosome pathway has been recently associated with the rapid degradation of approximately 50 % of pre-tRNA molecules (Gudipati et al. 2012). This pathway requires the action of the TRAMP (Trf4/Air2/Mtr4 polyadenylation) complex which, making use of its poly (A) polymerase activity, adds a poly (A) tail at the end of the tRNA to be degraded. Subsequently, this tRNA-TRAMP complex interacts with the nuclear exosome machinery leading to the tRNA degradation by 3′-5′ exonucleolytic activity (Parker 2012). Defects in maturation, modification, and in the 3′ end processing can be targeted by TRAMP and processed by the exosome (further discussed in section "Chemical modifications on tRNAs") (Parker 2012).

The RTD pathway involves the activity of 5′-3′ exonucleases located both in nucleus (the Rat1 exonuclease) and in cytoplasm (the Xrn1 exonuclease) (Chernyakov et al. 2008) (more details in Chap. 7). The Met22 protein, involved in the methionine biosynthetic pathway, also plays a role in the RTD pathway, thus connecting amino acid synthesis with tRNA turnover and quality control (Dichtl et al. 1997). This decay pathway seems to control the correct tertiary structure of certain tRNAs bearing modifications that stabilize their structure, since not all the tRNAs with the same modification are degraded by RTD (Whipple et al. 2011). The subcellular localization of the RTD pathway is still an open question, but it is likely that it takes place both in nucleus and in cytoplasm.

In addition to those exonucleolytic tRNA decay pathways, tRNAs can also be degraded by Rny1, an RNase T2 endonuclease. Rny1 is located in membrane-bound compartments, mainly vacuoles, and its activity produces cleavage at the anticodon loop of tRNAs in certain stress conditions (MacIntosh et al. 2001; Thompson et al. 2008; Thompson and Parker 2009). The activity of Rny1 may take place in the cytoplasm, by release of the enzyme during stress or directly in the vacuole, where the tRNA could be imported by some kind of autophagy process (Parker 2012). The function of these tRNA halves generated by endonucleolytic tRNA cleavage is not well understood but it has been proposed to play a role in the autophagy process (Thompson et al. 2008; Andersen and Collins 2012).

tRNA Transport

tRNAs are transcribed in the nucleus and function in the cytoplasm or the mitochondria, thus a nuclear export mechanism is mandatory and expected. However, studies on the process of tRNA trafficking have revealed many surprises. In addition to being exported out of the nucleus, tRNAs can be re-imported back to the nucleus and then exported again using specific transport mechanisms (Fig. 10.1).

The tRNA nuclear export through the nuclear pores is mediated by the small GTPase Ran. Ran regulates the nuclear export by its association with several proteins of the importin-β family. In yeast, those importin-β members are Los1 (Hopper et al. 1980; Hellmuth et al. 1998; Sarkar and Hopper 1998) and Msn5 (Takano et al. 2005; Shibata et al. 2006; Murthi et al. 2010). Los1 is thought to be the major tRNA exporter, albeit not the only one as demonstrated by the viability of Los1 deletion mutants (Hopper et al. 1980; Hurt et al. 1987). The interaction of Los1 with tRNA is strongly affected by the tRNA tertiary structure as well as the maturation of the 5′ and 3′ termini (Arts et al. 1998; Lipowsky et al. 1999; Cook et al. 2009; Lee et al. 2011). Therefore, Los1 works also as a quality control protein, preventing the export of misfolded or improperly processed tRNAs. Furthermore, under several stress conditions or in the absence of fermentable carbon sources, the subcellular localization of Los1 changes from nuclear to cytoplasmic. In the cytosol Los1 triggers several response mechanisms, thus linking cell homeostasis to nuclear tRNA export (Qiu et al. 2000; Ghavidel et al. 2007). The other importin-β, Msn5, plays a minor role in tRNA export in vertebrates (Bohnsack et al. 2002; Calado et al. 2002) but it is clear that, at least in yeast and *Drosophila*, Msn5 serves as a bona fide tRNA nuclear exporter (Takano et al. 2005; Shibata et al. 2006; Murthi et al. 2010). In these two organisms, Msn5 has been related to the export of intronless tRNAs and also with the reexport from nucleus to cytoplasm of tRNAs that previously reentered the nucleus by the tRNA retrograde mechanism (see below) (Murthi et al. 2010). Importantly, Los1 and Msn5 are likely not the only tRNA nuclear exporters since the double mutant for both proteins is not lethal (Takano et al. 2005).

Since tRNA splicing in yeast occurs in the cytoplasm, new questions arose when mature tRNAs were found in the nucleus under certain conditions such as nutrient deprivation or inhibition of tRNA aminoacylation (Hopper 2013). To explain this phenomenon, a tRNA retrograde import has been proposed (Yoshihisa et al. 2003; Stanford et al. 2004). The factors involved in such mechanism are not yet well understood, but some data suggest the participation of the Ssa2 heat shock protein in a Ran independent manner (Takano et al. 2005), and the Mtr10 importin-β in a Ran dependent manner (Shaheen and Hopper 2005). The tRNA imported by this retrograde mechanism under stress conditions can later return to the cytoplasm again by a mechanism named tRNA reexport (Whitney et al. 2007). Moreover, in normal conditions, tRNAs can be aminoacylated in the nucleus by a nuclear pool of aminoacyl-tRNA synthetases and these aminoacylated tRNAs may also use the tRNA reexport machinery to reach the cytoplasm (Hopper 2013). Notably, some evidence suggests that both Los1 and Msn5 may be participating in this tRNA reexport process (Murthi et al. 2010).

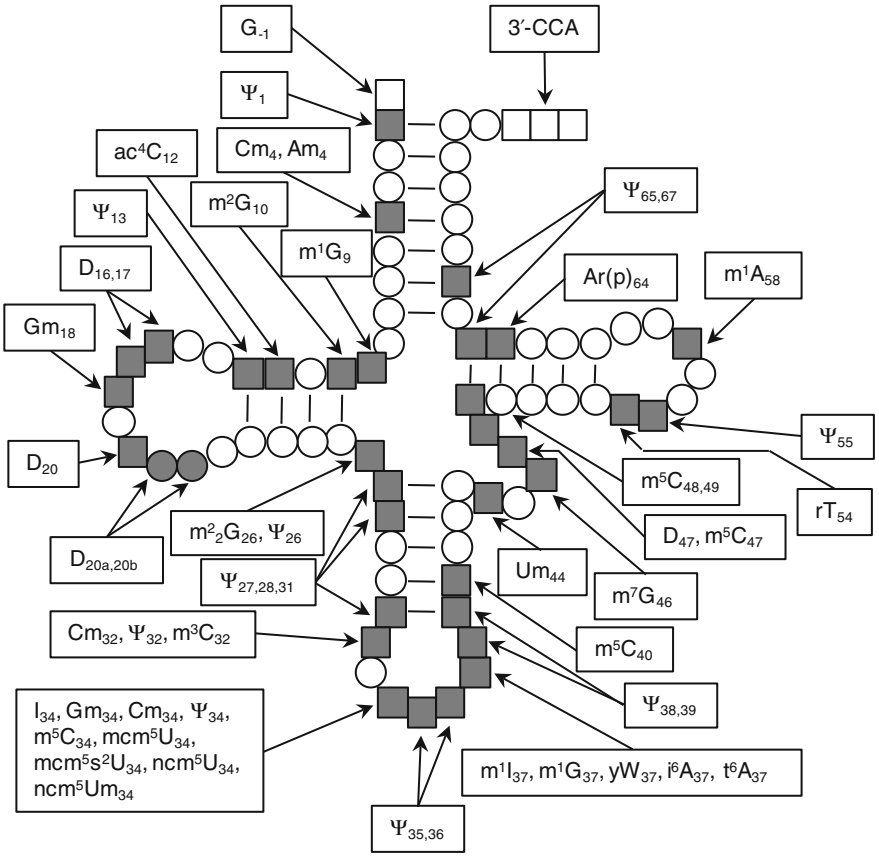
As mentioned above, some tRNAs can be exported into mitochondria. This has been observed for few nuclear encoded tRNAs in yeast (Rubio and Hopper 2011). Interestingly, yeast mitochondria encodes for the complete set of tRNAs necessary for the mitochondrial translation, so it has been proposed that those nuclear encoded tRNAs could be very important under stress situations. For example, yeast mitochondria use the $tRNA^{Lys}_{UUU}$ to read AAA and AAG codons and that capacity is achieved by a modification at the wobble position 34 of the tRNA. Under stress conditions, this tRNA is hypomodified and the nuclear encoded $tRNA^{Lys}_{UUU}$ seems to accomplish the task in the organelle (Kamenski et al. 2007).

Chemical Modifications on tRNAs

The primary sequence of tRNAs undergoes extensive posttranscriptional modifications. There are more than 60 different tRNA chemical modifications listed for eukarya (Cantara et al. 2011) and, for the majority of them, the enzymes responsible for such modifications have been characterized. The most detailed landscape on the chemistry and biology of tRNA modifications has been achieved in the budding yeast *S. cerevisiae*. In this section, we will use this organism as a model to describe the main tRNA modification enzymes (Fig. 10.3; Table 10.1) and their roles in biological processes.

Modifications Involving tRNA Methyltransferases

The most commonly found tRNA modifications are those catalyzed by tRNA methyltransferases (Trms). In *S. cerevisiae*, 18 different Trms have been described and these are sufficient for methylation of almost all known methylated tRNA residues (Towns and Begley 2012).



Universally conserved tRNA modifications

Cm, Um, Gm, D, Ψ, m⁵C, m⁵U, m³U, m¹G, m⁷G, m²G, m²₂G, m¹A, m⁶A, s²U, ac⁴C, t⁶A

Fig. 10.3 Schematic representation of the tRNA clover-leaf secondary structure. Modifications present in *Saccharomyces cerevisiae* are shown. *Circles* and *Squares* nucleosides. *Lines* base pairing. *Empty circles* unmodified residues. *Empty squares* posttranscriptional addition of nucleosides. *Filled squares* modified residues. *Filled circles* nucleosides at positions 20a and 20b present in some tRNAs (these residues can be found modified). The additional *box* with modifications shown in bold corresponds to those modifications universally conserved in all domains of life (but not necessarily present in *S. cerevisiae*). Abbreviations: G_{-1} (addition of guanine at position −1), *Ψ* (pseudouridine), *Cm/Am/Gm/Um* (2′O-methylcytidine/adenosine/guanosine/uridine, respectively), $m^1G/m^1A/m^1I$ (1-methylguanosine/adenosine/inosine, respectively), m^2G (2-methylguanosine), m^2_2G (N2, N2-dimethylguanosine), m^7G (7-methylguanosine), m^3C (3-methylcytidine), m^5C (5-methylcytidine), mcm^5U (5-methoxycarbonylmethyluridine), mcm^5s^2U (5-methoxycarbonylmethyl-2-thiouridine), ncm^5U (5-carbamoylmethyluridine), ncm^5Um (5-carbamoylmethyl-2′O-methyluridine), ac^4C (4-acetylcytidine), *D* (dihydrouridine), *I* (inosine), *yW* (wybutosine), i^6A (N6-isopentyladenosine), t^6A (N6-threonylcarbamoyladenosine), *rT* (ribothymidine), *Ar(p)* (2′O-ribosyladenosine phosphate). Underscript at the end of the modification indicates the affected residue(s)

Table 10.1 Genes encoding for enzymes that modify tRNA residues posttranscriptionally in *Saccharomyces cerevisiae*

Gene	Posttranscriptional modification	Position on tRNA	Target tRNA[a]	Essential/not essential gene[b]
CCA1	CCA	3′-end	All tRNAs	Essential
DUS1	D	16, 17	*Pos 16:* Ala(AGC), Arg(ACU), Arg(ACG), Asn(GUU), Asp(GUC), Cys(GCA), Gly(GCC), Gly(UCC), His(GUG), Ile(AAU), Ile(UAU), Leu(UAG), Leu(UAA), Lys(CUU), Lys(UUU), Met(CAU), Phe(GAA), Pro(UGG), Ser(CGA), Ser(UGA), Ser(AGA), Thr(AGU), Trp(CCA), Tyr(GUA), Val(AAC) *Pos 17:* Asn(GUU), Ile(AAU), Lys(CUU), Phe(GAA), Thr(AGU), Arg(UCU)	Not essential
DUS2	D	20	Ala(AGC), Arg(ACG), Asn(GUU), Asp(GUC), Cys(GCA), Gly(GCC), Gly(UCC), His(GUG), Ile(AAU), Ile(UAU), Leu(UAG), Leu(CAA), Leu(UAA), Pro(UGG), Ser(CGA), Ser(UGA), Ser(AGA), Thr(AGU), Trp(CCA), Tyr(GUA), Val(UAC), Val(AAC)	Not essential
DUS3	D	47	Ala(AGC), Arg(ACU), Arg(ACG), Asn(GUU), Cys(GCA), Ile(AAU), Ile(UAU), Lys(CUU), Met(CAU), Thr(AGU), Trp(CCA), Tyr(GUA), Val(UAC), Val(CAC), Val(AAC)	Not essential
DUS4	D	20a, 20b	Asn(GUU), Glu(CUC), Gly(UCC), His(GUG), Ile(AAU), Ile(UAU), Leu(UAG), Ser(CGA), Ser(UGA), Ser(AGA), Tyr(GUA), Val(UAC), Val(CAC), Val(AAC)	Not essential

(continued)

Table 10.1 (continued)

Gene	Posttranscriptional modification	Position on tRNA	Target tRNA[a]	Essential/not essential gene[b]
[c]IKI1, IKI3, ELP2-4, ELP6, KTI11-14, SIT4, SAP185, SAP190	"cm^5" modification on mcm^5U, mcm^5s^2U, ncm^5U and ncm^5Um	34	*$Cm^5(mcm^5U)$*: Arg(UCU), Gly(UCC) *$Cm^5(mcm^5s^2U)$*: Lys(UUU), Gln(UUG), Glu(UUC) *$Cm^5(ncm^5U)$*: Ala(UGC), Pro(UGG), Ser(UGA), Thr(UGU), Val(UAC) *$Cm^5(ncm^5Um)$*: Leu(UAA)	Only IKI1 essential
MOD5	I^6A	37	Cys(GCA); Ser(CGA); Ser(UGA); Ser(AGA); Tyr(GUA); Gly(UCC);	Not essential[e]
NFS1, ISU1, ISU2, CFD1, NBP35, CIA1, URM1, UBA4, NCS2, NCS6, TUM1	"s^2" modification on mcm^5s^2U	34	*NCS2 and NCS6*: Glu(UUC), Lys(UUU), Gln(UUG)	Not essential
PUS1	Ψ	1, 26, 27, 28, 34, 35, 36, 65, 67	*Pos 1*: Arg(ACG) *Pos 26*: Trp(CCA) *Pos 27*: Arg(ACU), Arg(ACG), Glu(CUC), Ile(UAU), Lys(CUU), Lys(UUU), Met(CAU), Trp(CCA), Val(UAC), Val(CAC) *Pos 28*: Lys(UUU), Trp(CCA), Val(CAC) *Pos 34*: Ile(UAU) *Pos 35*: Tyr(GUA) *Pos 36*: Ile(UAU) *Pos 65*: Trp(CCA) *Pos 67*: Ile(UAU), Lys(UUU)	Not essential

(continued)

Table 10.1 (continued)

Gene	Posttranscriptional modification	Position on tRNA	Target tRNA[a]	Essential/not essential gene[b]
PUS3	Ψ	38, 39	*Pos 38*: Ala(AGC), Gly(GCC), His(GUG), Thr(UAG) *Pos 39*: Arg(ACU), Arg(ACG), Arg(UCU), Asn(GUU), Cys(GCA), His(GUG), Leu(CAA), Leu(UAA), Lys(CUU), Met(CAU), Phe(GAA), Ser(AGA), Ser(CGA), Ser(UGA), Thr(AGU), Trp(CCA), Trp(UCA), Tyr(GUA)	Not essential[e]
PUS4	Ψ	55	Ala(AGC), Arg(ACU), Arg(ACG), Arg(UCU), Asn(GUU), Asp(GUC), Cys(GCA), Glu(CUC), Gly(GCC), Gly(UCC), His(GUG), Ile(AAU), Ile(GAU), Ile(UAU), Leu(UAG), Leu(UAA), Lys(CUU), Lys(UUU), Met(CAU), Phe(GAA), Pro(UGG), Ser(CGA), Ser(UGA), Ser(AGA), Ser(GCU), Thr(AGU), Thr(UAG), Trp(CCA), Trp(UCA), Tyr(GUA), Val(AAC), Val(CAC), Val(UAC)	Not essential
PUS6	Ψ	31	Met(CAU), Gly(UCC), Ile(GAU), Leu(UAA), Lys(UUU), Ser(GCU), Thr(UAG)	Not essential
PUS7	Ψ	13, 35	*Pos 13*: Asp(GUC), Glu(CUC), Gly(GCC), Gly(UCC), His(GUG), Pro(UGG), Val(CAC), Val(AAC) *Pos 35*: Tyr(GUA)	Not essential
PUS8	Ψ	32	Asp(GUC), Cys(GCA), Gly(GCC), His(GUG), Leu(UAG), Leu(CCA), Pro(UGG), Ser(AGA), Val(UAC), Val(CAC), Val(AAC)	Not essential
RIT1	Ar(p)	64	Initiator tRNA Met	Not essential
[d]SUA5	t^6A	37	All tRNAs decoding ANN codons bearing A37	Not essential[e]

(continued)

Table 10.1 (continued)

Gene	Posttranscriptional modification	Position on tRNA	Target tRNA[a]	Essential/not essential gene[b]
TAD1	I	37	Ala(AGC)	Not essential
TAD2, TAD3	I	34	Ala(AGC), Arg(ACG), Ile(AAU), Ser(AGA), Thr(AGU), Val(AAC), Pro(AGG)	Both genes essential
TAN1	ac^4C	12	Leu(UAG), Leu(CAA), Leu(UAA), Ser(CGA), Ser(UGA), Ser(AGA)	Not essential[e]
THG1	G	−1 (5′end)	His	Essential
TRM1	m^2_2G	26	Ala(AGC), Arg(ACG), Arg(ACU), Asn(GUU), Leu(CAA), Leu(UAA), Leu(UAG), Lys(CUU), Lys(UUU), Met(CAU), Phe(GAA), Pro(UGG), Ser(AGA), Ser(CGA), Ser(UGA), Thr(AGU), Thr(UAG), Tyr(GUA), Val(CAC)	Not essential
TRM2	m^5U	54	Arg(ACU), Arg(ACG), Asn(GUU), Asp(GUC), Cys(GCA), Glu(CUC), Gly(GCC), Gly(UCC), His(GUG), Ile(AAU), Ile(UAU), Leu(CAA), Leu(UAA), Leu(UAG), Lys(CUU), Lys(UUU), Met(CAU), Phe(GAA), Pro(UGG), Ser(CGA), Ser(UGA)	Not essential
TRM3	Gm	18	Leu(UAG), Leu(CAA), Leu(UAA), His(GUG), Tyr(GUA), Ser(AGA), Ser(CGA), Ser(UGA), Trp(CCA)	Not essential
TRM4	m^5C	34, 40, 48, 49	*Pos 34*: Leu(CUA) *Pos 40*: Phe(GAA) *Pos 48*: Asn(GUU), Cys(GCA), Ile(AAU), Ile(UAU), Leu(CAA), Leu(UAA), Leu(UAG), Lys(UUU), Met(CAU), Ser(AGA), Ser(CGA), Ser(UGA), Thr(AGU), Tyr(GUA) *Pos 49*: Arg(ACG), Asp(GUC), Glu(CUC), Gly(GCC), His(GUG), Phe(GAA), Val(AAC), Val(CAC), Val(UAC)	Not essential

(continued)

Table 10.1 (continued)

Gene	Posttranscriptional modification	Position on tRNA	Target tRNA[a]	Essential/not essential gene[b]
TRM5	m^1G, m^1I, yW	37	*m^1G*: Asp(GUC), His(GUG), Leu(CAA), Leu(UAA), Leu(UAG), Phe(GAA), Pro(UGG), Ser(UGA), Thr(UAG) *m^1I*: Ala(AGC) *yW*: Phe(GAA)	Not essential[e]
TRM6, TRM61	m^1A	58	Arg(ACG), Arg(ACU), Asn(GUU), Cys(GCA), Ile(AAU), Ile(UAU), Leu(UAA), Leu(UAG), Lys(CUU), Lys(UUU), Met(CAU), Phe(GAA), Pro(UGG), Thr(AGU), Trp(CCA), Tyr(GUA), Val(AAC), Val(CAC), Val(UAC)	Both genes essential
TRM7	Cm, Gm. "Um" modification on ncm^5Um	32(Cm), 34(Cm,Gm, Um)	*Cm Pos 32*: Leu(UAA), Phe(GAA), Trp(CCA) *Cm Pos 34*: Trp(CCA) *Gm*: Phe(GAA) *Um(ncm^5Um)*: Leu(UAA)	Not essential[e]
TRM8, TRM82	m^7G	46	Cys(GCA), Ile(UAU), Lys(CUU), Lys(UUU), Met(CAU), Phe(GAA), Pro(UGG), Trp(CCA), Val(AAC), Val(CAC)	Not essential
TRM9, TRM112	"m" modification on mcm^5U and mcm^5s^2U	34	*m(mcm^5U)*: Arg(UCU) *m(mcm^5s^2U)*: Glu(UUC)	Not essential[e]
TRM10	m^1G	9	Ala(AGC), Arg(ACG), Arg(UCU), Cys(GCA), Gly(CCC), Gly(GCC), Ile(AAU), Lys(CUU), Met(CAU), Pro(UGG), Thr(AGU), Thr(CGU), Trp(CCA), Val(AAC), Val(UAC)	Not essential
TRM11, TRM112	m^2G	10	Arg(ACG), Arg(ACU), Asn(GUU), Ile(AAU), Ile(UAU), Leu(CAA), Leu(UAA), Leu(UAG), Lys(CUU), Met(CAU), Phe(GAA), Thr(AGU), Trp(CCA), Tyr(GUA), Val(CAC), Val(UAC)	Not essential

(continued)

Table 10.1 (continued)

Gene	Posttranscriptional modification	Position on tRNA	Target tRNA[a]	Essential/not essential gene[b]
TRM13	Am, Cm	4	*Am*: His(GUG) *Cm*: Gly(GCC), Pro(UGG)	Not essential
TRM44	Um	44	Ser	Not essential
TRM140	m^3C	32	Ser(AGA), Thr(AGU)	Not essential
TYW1-4	yW	37	Phe(GAA)	Not essential

[a] All tRNA substrates (cytoplasmic or mitochondrial) that have been validated experimentally according to Modomics (Machnicka et al. 2013) are indicated
[b] Data obtained from the Saccharomyces Genome Database (Cherry et al. 2012). See also (Phizicky and Hopper 2010; Novoa and Ribas de Pouplana 2012)
[c] Components of the Elongator complex or associated to the Elongator complex
[d] Component of the ECK/KEOPS complex
[e] While these genes are not essential, yeast null mutants show several defects such as temperature sensitivity, slow growth, paromomycin sensitivity, or general sickness

Interestingly, while methylation of tRNA nucleotides is important for tRNA structure, function, and stability (see below), complete deletion of individual Trm genes in *S. cerevisiae* generally results in only mild phenotypes, or phenotypes under particular stress conditions. The exception to this is deletion of the Trm6/Trm61 complex, in charge of producing 1-methyladenosine (m^1A) at position 58 of many eukaryotic tRNAs, which was shown to be essential in budding yeast (Anderson et al. 2000). The deletion of *TRM5*, that catalyzes the formation of 1-methylguanosine (m^1G), 1-methylinosine (m^1I), and the first methylation step for synthesis of wybutosine (yW) (all at position 37 of the tRNA) results in dramatic growth defects but it is not an essential gene (Bjork et al. 2001).

An important observation regarding the methylation of tRNAs came from Chan and colleagues who showed that the methylation pattern for a given tRNA substrate can significantly change upon cell treatment with toxic chemical stimulants such as hydrogen peroxide (Chan et al. 2010). This is consistent with observed phenotypes in different yeast Trm null mutants when put under stress. Trm9 forms a complex with Trm112 to catalyze the last methylation step for 5-methoxycarbonylmethyluridine (mcm^5U) or 5-methoxycarbonylmethyl-2-thiouridine (mcm^5s^2U) at position 34 of different tRNAs. Deletion of *TRM9* in *S. cerevisiae* results in increased sensitivity to DNA damage (Begley et al. 2007) and to paromomycin, a translational inhibitor, when cells are grown at 37 °C (Kalhor and Clarke 2003). Likewise, deletion of *TRM4*, responsible for 5-mehtylcytidine (m^5C) modification, confers hypersensitivity to oxidative stress (Chan et al. 2012).

Elongator is a highly conserved protein complex composed of 6 proteins (Elp1-6 in yeast) that has been associated with numerous cellular processes such as histone acetylation, regulation of exocytosis, transcriptional elongation, and others (Phizicky and Hopper 2010). The Elongator complex is also required for the early steps in the modification of U34, therefore affecting many derived modifications bearing a 5-carboxymethyl core structure (Huang et al. 2005). In *S. pombe* lack of a functional Elongator complex affects mcm^5s^2U modification at position 34 of tRNAs and results in enhanced sensitivity to hydrogen peroxide (Fernandez-Vazquez et al. 2013).

Some yeast *TRM* deletions have severe consequences when combined with the deletion of additional tRNA modification enzymes. This is the case for mutants missing Trm8/Trm82, the enzymes responsible for formation of 7-methylguanosine (m^7G) at position 46 in tRNAs. *S. cerevisiae* lacking the *TRM8* and *TRM82* genes can grow normally under standard laboratory conditions, but are temperature-sensitive when grown in synthetic media containing glycerol (Alexandrov et al. 2005). However, under this deletion background, cells also deleted for other non-essential tRNA modification enzymes such as Trm4, Pus7, and Dus3, among others, showed distinct temperature-sensitive growth defects. Moreover, a mechanism for rapid tRNA degradation has been reported for endogenous $tRNA^{Val}_{ACC}$ in *S. cerevisiae* strains double-deleted for *TRM8* and *TRM4* (Alexandrov et al. 2006). Similarly, yeast strains lacking the nonessential enzymes Trm44 (in charge of 2′O-methylating uridine at position 44 in $tRNA^{Ser}$) and Tan1 (tRNA cytidine N4

acetyltransferase) show growth defects, while mutant strains for each single deletion do not. The growth phenotype under this double-deletion background, was attributed to reduced levels of $tRNA^{Ser}_{CGA}$ and $tRNA^{Ser}_{UGA}$ (Kotelawala et al. 2008), suggesting a role for tRNA modifications in tRNA stability.

Other tRNA Modifications

Beyond tRNA methyltransferases, there are several other tRNA modification enzymes that play crucial roles in tRNA biology. These include deamination enzymes, enzymes involved in thiolation, and acetyltransferases and isomerases, among others. Here, we will give a brief description of the most relevant ones and those that have been most widely studied.

Adenine (A) to Inosine (I) conversion is catalyzed by Adenosine Deaminases acting on tRNAs (ADATs), and occurs mainly at two different positions in tRNAs: the wobble position 34 and position 37 adjacent to the anticodon. In all eukaryotes, deamination of A37 is catalyzed by ADAT1 while deamination of A34 is catalyzed by the heterodimeric enzyme ADAT2/ADAT3. The homologues for these enzymes in yeast are named Tad1, Tad2, and Tad3, respectively. The roles for A-to-I modification vary with the position in the tRNA, with I37 likely influencing tRNA anticodon loop structure (Gerber et al. 1998; Grosjean et al. 1996) and I34 being more directly involved in codon-anticodon recognition (Crick 1966; Novoa et al. 2012). I37 is present only on eukaryotic $tRNA^{Ala}$ and is further modified to m^1I by an S-adenosyl-L-methionine-dependent methyltransferase. In *S. cerevisiae*, it was shown that Tad1 is capable to specifically modify A37 but not A34 and that *TAD1* null mutants were viable and showed no slow growth phenotype at 30 °C (Gerber et al. 1998).

While adenine can only pair efficiently with uridine; inosine can pair with uridine, adenine, and cytosine. This has raised the possibility that I34 plays a crucial role in expanding the number of codons a single tRNA can recognize. Indeed, very recently a role in shaping the genome structure and codon usage has been suggested for ADAT2(Tad2)/ADAT3(Tad3) in eukaryotes as an improved correlation between tRNA gene copy number and codon usage could be achieved when the effect of the I34 modification was considered in the analysis (Novoa et al. 2012). Consistent with this critical functional role for this modification, both *TAD2* and *TAD3* genes were shown to be essential in *S. cerevisiae* (Gerber and Keller 1999).

Thiolation is another very important modification that is found at U34 of several tRNA species resulting in 2-thiouridine (s^2U) (Bjork et al. 2007). In *S. cerevisiae*, this modification is catalyzed by the complex Ncs2/Ncs6 for cytosolic tRNAs and by Mtu1 for mitochondrial tRNAs. Following thiolation, U34 is usually further modified at the C5-position of the pyrimidine ring with other chemical groups. The combination of all of these modifications at U34 may have consequences for aminoacylation, translation efficiency and fidelity (Jackman and Alfonzo 2013).

Wybutosine (yW37) is present at position 37 of $tRNA^{Phe}$. It is a guanosine derivative resulting from several chemical modification steps. In *S. cerevisiae*, the reactions are catalyzed by 4 different enzymes Tyw1, Tyw2, Tyw3, and Tyw4 that produce a methyl imidazole ring, addition of an α-amino-α-carboxypropyl group, methylation of guanosine N3, and methylation of the α-carboxy group and methoxycarboxylation of the α-amino group; respectively. While this modification seems to be important for reading frame maintenance, none of the enzymes responsible for yW37 were reported to be essential in yeast (Phizicky and Hopper 2010).

Isomerization occurs for uridine at several positions in tRNAs to give rise to pseudouridine. This reaction is catalyzed by tRNA pseudouridine synthases (Pus). In *S. cerevisiae*, there are 6 different Pus that act on different tRNA substrates and different tRNA residues (Phizicky and Hopper 2010). Disruption of *DEG-1*, the gene encoding for Pus3 that modifies uridine at position 38 and 39 in yeast, was shown to reduce growth rate, especially at 37 °C (Lecointe et al. 1998).

N6-threonylcarbamoyl adenosine (t^6A) modification at position 37 of tRNAs is one of the few modifications at the anticodon loop that is present in all kingdoms of life. It is a complex modification involving several enzymes out of which some remain unknown for eukaryotes and archea (El Yacoubi et al. 2012). In *S. cerevisiae*, the gene *SUA5* responsible for the t^6A modification was shown to be dispensable; however, null strains for this gene showed severe growth phenotypes (El Yacoubi et al. 2009).

Role of tRNA Modifications in Structure, Function, and Stability of tRNAs

The role of tRNA modifications in tRNA structure and function has been recently reviewed (Jackman and Alfonzo 2013). Notably, while tRNAs are heavily modified by many different enzymes involving several types of chemical reactions, the overall result for such modifications in tRNA function seems to be rather subtle. Here, we will describe some of the most visible effects that tRNA modifications can have on tRNA structure, function, and stability.

Modifications in the region of the anticodon loop can affect protein translation. In general, modifications at position 37 sterically block Watson-Crick pairing and can influence frameshifting. For example, isopentenyl adenosine (i^6A) modification at position 37 in *S. pombe* was shown to increase the affinity of tRNAs for cognate codons while decreasing affinity for noncognate codons (Lamichhane et al. 2013). Likewise, the presence of m^1G at position 37 was shown to prevent frameshifting in *Salmonella typhimurium* (Bjork et al. 1989) and was likely the cause for the observed growth phenotypes upon deletion of *TRM5* in *S. cerevisiae* (Bjork et al. 2001). Similarly, Urbonavicius and colleagues showed that Trm5 methylation was necessary for proper in-frame translation affecting +1 frameshifting but not −1 frameshifting (Urbonavicius et al. 2001, 2003). Wybutosine 37, like other purine bases at position 37, helps stabilize codon-anticodon pairing by providing base stacking interactions within the A site of the ribosome (Konevega et al. 2004) and has been

shown to prevent −1 frameshifts in yeast (Waas et al. 2007). Finally, translation initiation defects have been reported for cytoplasmic tRNAs lacking the t^6A37 modification (Lin et al. 2010; El Yacoubi et al. 2011).

Modifications at position 34 of the tRNA can also affect translational fidelity; one such example in *S. cerevisiae* is the role of Trm9 modifications in a particular set of transcripts where it prevents protein synthesis errors and the triggering of protein stress response pathways (Patil et al. 2012). However, modifications at position 34 are considered more important for decoding than for misreading, as exemplified by the I34 modification mentioned above and by the thiolation of uridine (and posterior modifications) recently reviewed by Phizicky and Hopper (Phizicky and Hopper 2010).

Some modifications play key roles in tRNA structure and stability. These modifications are usually located at the core of the tRNA and either rigidify or increase flexibility of its structure. For example, pseudouridines keep their ribose rings in the strong binding C3′endo conformation, while dihydrouridines keep the sugar pucker in a DNA-like C2′endo conformation (Giege et al. 2012). The C3′endo sugar conformation can also protect the RNA from endogenous nucleases. Thermostability of tRNAs due to particular modifications has been well documented in bacteria, where some of these modifications are exclusive of thermophilic species (Motorin and Helm 2010); it remains to be explored whether novel thermostable tRNA modifications are present in thermophilic fungi. In this respect an in vitro comparison between a tRNA from *Humicola lanuginosa* and *Candida utilis* did not show major differences in melting temperatures; however, the tRNA from *H. lanuginosa* was more resistant to RNase A treatment, suggesting a more compact tRNA tertiary structure (Joshi and Cherayil 1987). Finally, modifications can also promote ion binding, as is the case of m^5C in yeast $tRNA^{Phe}$, which helps binding of magnesium at a distal site in the tRNA, thus promoting magnesium dependent conformational shifts (Chen et al. 1993).

Some modifications are critical to prevent the tRNA from entering certain degradation pathways. This has been exemplified previously when discussing *TRM8/TRM4* double mutants (Alexandrov et al. 2006; see above). Another example is the role for Trm2 in *S. cerevisiae* in stabilizing $tRNA^{Ser}_{CGA}$; although in this case the need for m^5U at position 54 did not seem to be the main factor affecting tRNA stability (Johansson and Bystrom 2002). As mentioned before, yeast $tRNA^{Ser}_{CGA}$ and $tRNA^{Ser}_{UGA}$ can be destabilized in the absence of the enzymes responsible for methyluridine at position 44 (Trm44) and N4-acetylcytidine at position 12 (Tan1) (Kotelawala et al. 2008). Finally, in *S. cerevisiae*, aberrant pre-$tRNA^{iMet}$ transcripts lacking (m^1A) at position 58 are recognized by the TRAMP complex (discussed above) and targeted for degradation (Phizicky and Hopper 2010).

There are also examples of modifications that serve as identity elements for the tRNA. We have already discussed the role of G_{-1} on amino acid charging by HisRS (Rudinger et al. 1994). Also, m^1G at position 37 of $tRNA^{Asp}$ has been described as an antideterminant for ArgRS (Putz et al. 1994). Lastly, the 2′O-

ribosyl phosphate modification at position 64 of the initiator tRNA in *S. cerevisiae*, catalized by Rit1, discriminates the initiator tRNAs from the elongator tRNAs during protein synthesis (Astrom and Bystrom 1994).

Mounting evidence now suggests that some tRNA modifications may have several roles and be linked to complex phenotypes and even human diseases (Torres et al. 2014). While some modifications may affect several tRNAs and general protein translation; others may affect single tRNAs and the reading of specific codons, leading to translation defects on specific proteins enriched in such codons (Novoa and Ribas de Pouplana 2012). For example, in yeast exposed to hydrogen peroxide, an increase in m^5C modification at the wobble position by Trm4 was observed in $tRNA^{Leu}_{CAA}$ leading to selective translation of mRNAs enriched in the TTG codon (Chan et al. 2012). On the other hand, general translational defects might be the cause of the multiple phenotypes associated with the Elongator complex. This complex is involved in the formation of mcm^5s^2U, and has been associated with different roles such as acetylation of histones H3 and H4, transcriptional elongation, regulation of exocytosis, filamentous growth signaling, and telomeric gene silencing (Phizicky and Hopper 2010; El Yacoubi et al. 2012). Similarly, the KEOPS (Kinase, Endopeptidase, and Other Small Proteins)/ECK (Endopeptidase-like kinase associated to transcribed chromatin) complex has been associated with multiple cellular functions such as transcription, telomere maintenance, and chromosome segregation and have recently been shown to be required for formation of t^6A34 modification (Srinivasan et al. 2011).

tRNA Modifications and Evolution

As discussed, the complete set of modifications of any given tRNA is involved in three different aspects of tRNA function: folding, molecular recognition by proteins or RNAs, and codon-anticodon pairing. Out of the more than a hundred possible modifications only 17 are universally present in all domains of life (Grosjean 2009) (Fig. 10.3), and a clear tendency in evolution to increase the number and complexity of tRNA modifications can be seen from Archaea to bacteria, and to eukarya (Jackman and Alfonzo 2013).

The modifications that play a role in tRNA folding are thought to influence subtle features of the structure of tRNA, rather than having a major impact on the folding of the molecule. Crystallographic studies comparing an unmodified $tRNA^{Phe}$ from *E. coli* to the fully modified $tRNA^{Phe}$ from yeast revealed only slight structural changes, with no effect, for example, on the anticodon conformation (Byrne et al. 2010). Structural modifications are generally thought to contribute to the rigidity or flexibility of the tRNA molecule (Giege et al. 2012), and tend to be conserved in organisms that can particularly benefit from such features as those living in extreme temperature conditions (Edmonds et al. 1991; McCloskey et al. 2001).

Recognition by enzymes and RNAs involved in the maturation, aminoacylation, or transit through the ribosome of the tRNA can also require specific modifications. A key molecular recognition of tRNAs takes place during their aminoacylation by aminoacyl-tRNA synthetases (ARS). This activity translates the genetic code from triplet information to amino acid side chain, and constitutes the main check point to maintain the fidelity of the whole process of gene translation. Indeed, some unmodified tRNA sequences would be too similar to be easily distinguished by ARSs directly, and depend on modifications to prevent misacylation. An example of this situation is the bacterial $tRNA^{Ile}_{CAU}$. This tRNA can be easily misacetylated by methionyl-tRNA synthetase (MRS), but this reaction is prevented by the modification of its C34 into lysidine (L) by lysidine-tRNA synthase (TilS) (Muramatsu et al. 1988; Soma et al. 2003). Interestingly, many of these identity marks were established late in evolution and are not conserved among species. For example, in the bacteria *Mycoplasma penetrans* L34 is not required for MRS rejection of $tRNA^{Ile}_{CAU}$ which, instead, depends upon interactions with the acceptor stem (Jones et al. 2008, 2013).

Interestingly, tRNA misacylation is a physiological feature of some yeast tRNAs, which can be charged with two amino acids by different ARS (Moura et al. 2010). Indeed, a *C. albicans* $tRNA^{Ser}_{GAC}$ is charged normally with serine by seryl-tRNA synthetase (SRS), but is also charged with leucine by leucyl-tRNA synthetase (LRS) (Moura et al. 2010). It should be noted that this phenomenon results in a statistical proteome where, under normal circumstances, 3 % of all coded serines are replaced by leucines in a random manner. Remarkably the levels of mistranslation can increase dramatically as a response to environmental factors, resulting in important changes in cell adhesion and colony structure (Silva et al. 2007). *C. albicans* has evolved to tolerate this adaptive strategy, and the vast majority of CUG codons in its genome code for surface amino acids in order to minimize the effect of this decoding flexibility upon protein structure (Rocha et al. 2011).

Base modifications in the anticodon loop are of utmost importance due to their potential influence in codon-anticodon recognition. In fact, the most intensively modified part of the tRNA is the anticodon loop and, within it, the most often altered positions are positions 34 and 37 (Chan and Lowe 2009). Indeed, in the genome of the bacterial endosymbiont Candidatus, *Riesia pediculicola*, which is characterized by its extreme gene reduction, the only tRNA modification enzymes conserved are those responsible for modifications at positions 34, 37, 38, and 39 (de Crecy-Lagard et al. 2012). The tRNA wobble nucleotide is usually modified and such modifications are essential for tRNA decoding. Interestingly, when looking at the three domains of life, it is clear that not all modifications are universal. For example, Cm, s^2U, Gm, and Um are found in eukarya, bacteria, and archea, but s^2Um is present only in archea, mcm^5s^2U is present only in eukarya, and $cmnm^5s^2Um$ is present only in bacteria (Grosjean et al. 2010). This implies that different species have developed different decoding strategies for some anticodon-codon pairings.

The relatively good correlation between tRNA gene content and tRNA abundance implies that, to optimize translation, a co-evolution probably exists between tRNA gene content in the genome and the activities of the enzymes that modify the anticodon loop. A balance is necessary between tRNA abundance and the introduced anticodon modifications to ensure proper translation of the degenerate genetic code in a given codon usage context. Studies from Grosjean and coworkers have proposed different decoding strategies to explain the genomic composition of tRNAs genes based on the sequences and tRNA modifications specifically at the wobble position (Grosjean et al. 2010). Interestingly, in all the organisms checked, an A34NN anticodon never coexists with a G34NN anticodon decoding for the same amino acid. In Archaea, only the G34NN are used to read the pyrimidine-ending codons (G at position 34 of the tRNA and C or U at the third position of the codon). In bacteria the same usually applies, except for the case of $tRNA^{Arg}$, using an A34NN anticodon instead of a G34NN. The A34 in this anticodon is modified to inosine (I) by the homodimeric TadA enzyme, specific for $tRNA^{Arg}_{ACG}$, and thus can perform I34:C3, I34:U3, and less efficiently I34:A3 wobble pairings (Murphy and Ramakrishnan 2004; Wolf et al. 2002).

Other unmodified A34NN tRNAs in bacteria such as $tRNA^{Thr}$ from *M. capricolum*, or $tRNA^{Leu}$ from *M. synovia*, are very rare and are used to decode the four synonymous codons of their boxes (Lim 1995; Grosjean et al. 2010). On the other hand, in eukarya an A34 is systematically used instead of a G34 for the four box tRNA sets, except in the case of $tRNA^{Gly}$. All of these A34 are also modified to I, but in this case by the heterodimeric complex Tad2/Tad3 in yeast and ADAT2/ADAT3 in other eukaryotes as mammals (Gerber and Keller 1999). We have recently proposed that the evolution of modification enzymes acting at this wobble position played a central role in the difference in tRNA gene content, and the codon usage, of bacteria and eukarya (Novoa et al. 2012). Moreover, we showed that in these species, a good correlation between tRNA gene copy number and codon usage can only be found when the effect of these modifications, catalyzed by hetADATs (I34 modification) in eukarya and by uridine methyltransferases in bacteria, are taken into account (Novoa et al. 2012).

While there are different decoding strategies in the different kingdoms, some decoding strategies are universally conserved. This is the case, for example, for the preferential usage of U_{34}NN tRNAs for decoding synonymous codons on amino acids in the 4 box set; or the use of U_{34}NN or G_{34}NN tRNAs to decode the purine-ending or pyrimidine-ending codons, respectively, on amino acids in the 2 box split set (Grosjean et al. 2010). In this latter case, wobble modifications are usually required and are catalyzed by phylogenetically unrelated enzymes in different organisms. An interesting example is that of s^2U34 modification. This modification is universally conserved but the enzymes responsible for its synthesis evolved independently in bacteria and eukaryotes. Extant eukaryotes, however, contain both a bacterial- and an eukaryal modification system that act, respectively, in the mitochondria and in the nucleus (Jackman and Alfonzo 2013).

Several examples exist of convergent evolution involving tRNA modification enzymes. For instance, the chemical modification m^1G37 is highly conserved in bacteria, eukarya, and archea; however, it is catalyzed by TrmD (in bacteria) and by Trm5 (in eukarya and archea), two evolutionary unrelated enzymes (Christian and Hou 2007). A more extreme case is that of the t^6A37 modification which is also universally conserved in all kingdoms but the synthesis pathways differ dramatically among different species. On one hand, a set of core enzymes are conserved in all organisms for the synthesis of t^6A37; and on the other hand, specific kingdom-specific or even organelle-specific enzymes are required to complete the synthesis pathway for this modification (El Yacoubi et al. 2012).

Just as different enzymatic solutions exist for the same modification, members of the same enzyme family may catalyze different modification reactions in different organisms. Such is the case of the RlmD family, which in *E. coli* catalyze the formation of m^5U in the 23S rRNA (Madsen et al. 2003), but in archea is responsible for the formation of m^5U in tRNAs (Urbonavicius et al. 2008). This is just one of several cases for divergent functional evolution for tRNA modification enzymes. This phenomena limits our ability to predict the function of homologous enzymes from different organisms; and the experimental validation of each enzyme candidate for the catalysis of a given tRNA modification is crucial. Moreover, as we move on to more complex organisms, the number of co-existing genes coding for potential tRNA modification enzymes tends to increase. For example, humans have more Trm homologues than yeast (Towns and Begley 2012), possibly because such enzymes in humans have a different set of substrates.

The reason why some tRNA modifications are conserved while others are not remains unclear. While there are 61 codons for decoding all 20 proteinogenic amino acids, not all possible 61 tRNA species have been found in any organism. As explained above, different species resolve this issue by relying on different tRNA anticodon modifications, especially that of position 34, that allow tRNAs to recognize more than one codon. It has been proposed (Grosjean et al. 2010) that the lack of some tRNA species could be a trait from the Last Universal Common Ancestor (LUCA). According to this theory, LUCA would have only incorporated a subset of the 20 common proteinogenic amino acids to its genetic code. Therefore, the decoding capacity of its tRNAs did not rely on position 34 of the anticodon. After the split of the three domains, organisms gradually incorporated new amino acids to the genetic code, and that lead to the split of several four codon boxes into 2/2 or 3/1 alternatives. To maintain a faithful and efficient translation, these organisms evolved different decoding strategies using tRNA anticodon modifications to fine tune the decoding capacity of their tRNAs. It should be noted that this evolutionary scenario is also possible if one considers LUCA to have incorporated all the 20 amino acids, but with a certain degree of inefficiency or translation error. Posterior and independent solutions would be found in each domain of life to improve the efficiency and accuracy of the system.

References

Acker J, Conesa C, Lefebvre O (2013) Yeast RNA polymerase III transcription factors and effectors. Biochim Biophys Acta 1829:283–295

Aebi M, Kirchner G, Chen JY, Vijayraghavan U, Jacobson A, Martin NC, Abelson J (1990) Isolation of a temperature-sensitive mutant with an altered tRNA nucleotidyltransferase and cloning of the gene encoding tRNA nucleotidyltransferase in the yeast Saccharomyces cerevisiae. J Biol Chem 265:16216–16220

Alexandrov A, Chernyakov I, Gu W, Hiley SL, Hughes TR, Grayhack EJ, Phizicky EM (2006) Rapid tRNA decay can result from lack of nonessential modifications. Mol Cell 21:87–96

Alexandrov A, Grayhack EJ, Phizicky EM (2005) tRNA m7G methyltransferase Trm8p/Trm82p: evidence linking activity to a growth phenotype and implicating Trm82p in maintaining levels of active Trm8p. RNA 11:821–830

Andersen KL, Collins K (2012) Several RNase T2 enzymes function in induced tRNA and rRNA turnover in the ciliate Tetrahymena. Mol Biol Cell 23:36–44

Anderson J, Phan L, Cuesta R, Carlson BA, Pak M, Asano K, Bjork GR, Tamame M, Hinnebusch AG (1998) The essential Gcd10p-Gcd14p nuclear complex is required for 1-methyladenosine modification and maturation of initiator methionyl-tRNA. Genes Dev 12:3650–3662

Anderson J, Phan L, Hinnebusch AG (2000) The Gcd10p/Gcd14p complex is the essential two-subunit tRNA(1-methyladenosine) methyltransferase of Saccharomyces cerevisiae. Proc Natl Acad Sci USA 97:5173–5178

Arimbasseri AG, Rijal K, Maraia RJ (2013) Transcription termination by the eukaryotic RNA polymerase III. Biochim Biophys Acta 1829:318–330

Arts GJ, Kuersten S, Romby P, Ehresmann B, Mattaj IW (1998) The role of exportin-t in selective nuclear export of mature tRNAs. EMBO J 17:7430–7441

Astrom SU, Bystrom AS (1994) Rit1, a tRNA backbone-modifying enzyme that mediates initiator and elongator tRNA discrimination. Cell 79:535–546

Begley U, Dyavaiah M, Patil A, Rooney JP, Direnzo D, Young CM, Conklin DS, Zitomer RS, Begley TJ (2007) Trm9-catalyzed tRNA modifications link translation to the DNA damage response. Mol Cell 28:860–870

Bjork GR, Huang B, Persson OP, Bystrom AS (2007) A conserved modified wobble nucleoside (mcm5s2U) in lysyl-tRNA is required for viability in yeast. RNA 13:1245–1255

Bjork GR, Jacobsson K, Nilsson K, Johansson MJ, Bystrom AS, Persson OP (2001) A primordial tRNA modification required for the evolution of life? EMBO J 20:231–239

Bjork GR, Wikstrom PM, Bystrom AS (1989) Prevention of translational frameshifting by the modified nucleoside 1-methylguanosine. Science 244:986–989

Boguta M, Czerska K, Zoladek T (1997) Mutation in a new gene MAF1 affects tRNA suppressor efficiency in Saccharomyces cerevisiae. Gene 185:291–296

Bohnsack MT, Regener K, Schwappach B, Saffrich R, Paraskeva E, Hartmann E, Gorlich D (2002) Exp5 exports eEF1A via tRNA from nuclei and synergizes with other transport pathways to confine translation to the cytoplasm. EMBO J 21:6205–6215

Brogaard K, Xi L, Wang JP, Widom J (2012) A map of nucleosome positions in yeast at base-pair resolution. Nature 486:496–501

Byrne RT, Konevega AL, Rodnina MV, Antson AA (2010) The crystal structure of unmodified tRNAPhe from Escherichia coli. Nucleic Acids Res 38:4154–4162

Calado A, Treichel N, Muller EC, Otto A, Kutay U (2002) Exportin-5-mediated nuclear export of eukaryotic elongation factor 1A and tRNA. EMBO J 21:6216–6224

Cantara WA, Crain PF, Rozenski J, McCloskey JA, Harris KA, Zhang X, Vendeix FA, Fabris D, Agris PF (2011) The RNA modification database, RNAMDB: 2011 update. Nucleic Acids Res 39:D195–D201

Carles C, Treich I, Bouet F, Riva M, Sentenac A (1991) Two additional common subunits, ABC10 alpha and ABC10 beta, are shared by yeast RNA polymerases. J Biol Chem 266:24092–24096

Cook AG, Fukuhara N, Jinek M, Conti E (2009) Structures of the tRNA export factor in the nuclear and cytosolic states. Nature 461:60–65

Copela LA, Fernandez CF, Sherrer RL, Wolin SL (2008) Competition between the Rex1 exonuclease and the La protein affects both Trf4p-mediated RNA quality control and pre-tRNA maturation. RNA 14:1214–1227

Coughlin DJ, Pleiss JA, Walker SC, Whitworth GB, Engelke DR (2008) Genome-wide search for yeast RNase P substrates reveals role in maturation of intron-encoded box C/D small nucleolar RNAs. Proc Natl Acad Sci USA 105:12218–12223

Crick FH (1963) On the genetic code. Science 139:461–464

Crick FH (1966) Codon–anticodon pairing: the wobble hypothesis. J Mol Biol 19:548–555

Chan CT, Dyavaiah M, Demott MS, Taghizadeh K, Dedon PC, Begley TJ (2010) A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. PLoS Genet 6:e1001247

Chan CT, Pang YL, Deng W, Babu IR, Dyavaiah M, Begley TJ, Dedon PC (2012) Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. Nat Commun 3:937

Chan PP, Lowe TM (2009) GtRNAdb: a database of transfer RNA genes detected in genomic sequence. Nucleic Acids Res 37:D93–D97

Chen Y, Beck A, Davenport C, Shattuck D, Tavtigian SV (2005) Characterization of TRZ1, a yeast homolog of the human candidate prostate cancer susceptibility gene ELAC2 encoding tRNase Z. BMC Mol Biol 6:12

Chen Y, Sierzputowska-Gracz H, Guenther R, Everett K, Agris PF (1993) 5-Methylcytidine is required for cooperative binding of Mg2+ and a conformational transition at the anticodon stem-loop of yeast phenylalanine tRNA. Biochemistry 32:10249–10253

Chernyakov I, Whipple JM, Kotelawala L, Grayhack EJ, Phizicky EM (2008) Degradation of several hypomodified mature tRNA species in Saccharomyces cerevisiae is mediated by Met22 and the 5′-3′ exonucleases Rat1 and Xrn1. Genes Dev 22:1369–1380

Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, Christie KR, Costanzo MC, Dwight SS, Engel SR, Fisk DG, Hirschman JE, Hitz BC, Karra K, Krieger CJ, Miyasato SR, Nash RS, Park J, Skrzypek MS, Simison M, Weng S, Wong ED (2012) Saccharomyces genome database: the genomics resource of budding yeast. Nucleic Acids Res 40:D700–D705

Christian T, Hou YM (2007) Distinct determinants of tRNA recognition by the TrmD and Trm5 methyl transferases. J Mol Biol 373:623–632

Daoud R, Forget L, Lang BF (2012) Yeast mitochondrial RNase P, RNase Z and the RNA degradosome are part of a stable supercomplex. Nucleic Acids Res 40:1728–1736

De Crecy-Lagard V, Marck C, Grosjean H (2012) Decoding in Candidatus Riesia pediculicola, close to a minimal tRNA modification set? Trends Cell Mol Biol 7:11–34

Dhungel N, Hopper AK (2012) Beyond tRNA cleavage: novel essential function for yeast tRNA splicing endonuclease unrelated to tRNA processing. Genes Dev 26:503–514

Dichtl B, Stevens A, Tollervey D (1997) Lithium toxicity in yeast is due to the inhibition of RNA processing enzymes. EMBO J 16:7184–7195

Dieci G, Bosio MC, Fermi B, Ferrari R (2013) Transcription reinitiation by RNA polymerase III. Biochim Biophys Acta 1829:331–341

Dieci G, Fiorino G, Castelnuovo M, Teichmann M, Pagano A (2007) The expanding RNA polymerase III transcriptome. Trends Genet 23:614–622

Dumay-Odelot H, Durrieu-Gaillard S, da Silva D, Roeder RG, Teichmann M (2010) Cell growth- and differentiation-dependent regulation of RNA polymerase III transcription. Cell Cycle 9:3687–3699

Edmonds CG, Crain PF, Gupta R, Hashizume T, Hocart CH, Kowalak JA, Pomerantz SC, Stetter KO, McCloskey JA (1991) Posttranscriptional modification of tRNA in thermophilic archaea (Archaebacteria). J Bacteriol 173:3138–3148

El Yacoubi B, Bailly M, De Crecy-Lagard V (2012) Biosynthesis and function of posttranscriptional modifications of transfer RNAs. Annu Rev Genet 46:69–95

El Yacoubi B, Hatin I, Deutsch C, Kahveci T, Rousset JP, Iwata-Reuyl D, Murzin AG, De Crecy-Lagard V (2011) A role for the universal Kae1/Qri7/YgjD (COG0533) family in tRNA modification. EMBO J 30:882–893

El Yacoubi B, Lyons B, Cruz Y, Reddy R, Nordin B, Agnelli F, Williamson JR, Schimmel P, Swairjo MA, De Crecy-Lagard V (2009) The universal YrdC/Sua5 family is required for the formation of threonylcarbamoyladenosine in tRNA. Nucleic Acids Res 37:2894–2909

Ellis JC, Brown JW (2009) The RNase P family. RNA Biol 6:362–369

Fernandez-Vazquez J, Vargas-Perez I, Sanso M, Buhne K, Carmona M, Paulo E, Hermand D, Rodriguez-Gabriel M, Ayte J, Leidel S, Hidalgo E (2013) Modification of tRNA(Lys) UUU by elongator is essential for efficient translation of stress mRNAs. PLoS Genet 9:e1003647

Ge H, Roeder RG (1994) Purification, cloning, and characterization of a human coactivator, PC4, that mediates transcriptional activation of class II genes. Cell 78:513–523

Geiduschek EP, Kassavetis GA (2001) The RNA polymerase III transcription apparatus. J Mol Biol 310:1–26

Gerber A, Grosjean H, Melcher T, Keller W (1998) Tad1p, a yeast tRNA-specific adenosine deaminase, is related to the mammalian pre-mRNA editing enzymes ADAR1 and ADAR2. EMBO J 17:4780–4789

Gerber AP, Keller W (1999) An adenosine deaminase that generates inosine at the wobble position of tRNAs. Science 286:1146–1149

Ghavidel A, Kislinger T, Pogoutse O, Sopko R, Jurisica I, Emili A (2007) Impaired tRNA nuclear export links DNA damage and cell-cycle checkpoint. Cell 131:915–926

Giege R, Juhling F, Putz J, Stadler P, Sauter C, Florentz C (2012) Structure of transfer RNAs: similarity and variability. Wiley Interdisc Rev RNA 3:37–61

Grosjean H (2009) DNA and RNA modification enzymes: structure, mechanism, function and evolution. Landes Bioscience, Austin

Grosjean H, Auxilien S, Constantinesco F, Simon C, Corda Y, Becker HF, Foiret D, Morin A, Jin YX, Fournier M, Fourrey JL (1996) Enzymatic conversion of adenosine to inosine and to N1-methylinosine in transfer RNAs: a review. Biochimie 78:488–501

Grosjean H, De Crecy-Lagard V, Marck C (2010) Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes. FEBS Lett 584:252–264

Gu W, Hurto RL, Hopper AK, Grayhack EJ, Phizicky EM (2005) Depletion of Saccharomyces cerevisiae tRNA(His) guanylyltransferase Thg1p leads to uncharged tRNAHis with additional m(5)C. Mol Cell Biol 25:8191–8201

Gu W, Jackman JE, Lohan AJ, Gray MW, Phizicky EM (2003) tRNAHis maturation: an essential yeast protein catalyzes addition of a guanine nucleotide to the 5′ end of tRNAHis. Genes Dev 17:2889–2901

Gudipati RK, Xu Z, Lebreton A, Seraphin B, Steinmetz LM, Jacquier A, Libri D (2012) Extensive degradation of RNA precursors by the exosome in wild-type cells. Mol Cell 48:409–421

Gutmann B, Gobert A, Giege P (2012) PRORP proteins support RNase P activity in both organelles and the nucleus in Arabidopsis. Genes Dev 26:1022–1027

Hamada M, Huang Y, Lowe TM, Maraia RJ (2001) Widespread use of TATA elements in the core promoters for RNA polymerases III, II, and I in fission yeast. Mol Cell Biol 21:6870–6881

Hellmuth K, Lau DM, Bischoff FR, Kunzler M, Hurt E, Simos G (1998) Yeast Los1p has properties of an exportin-like nucleocytoplasmic transport factor for tRNA. Mol Cell Biol 18:6374–6386

Hernandez N (1993) TBP, a universal eukaryotic transcription factor? Genes Dev 7:1291–1308

Holzmann J, Frank P, Loffler E, Bennett KL, Gerner C, Rossmanith W (2008) RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. Cell 135:462–474

Hopper AK (2013) Transfer RNA post-transcriptional processing, turnover, and subcellular dynamics in the yeast Saccharomyces cerevisiae. Genetics 194:43–67

Hopper AK, Banks F (1978) A yeast mutant which accumulates precursor tRNAs. Cell 14:211–219
Hopper AK, Phizicky EM (2003) tRNA transfers to the limelight. Genes Dev 17:162–180
Hopper AK, Schultz LD, Shapiro RA (1980) Processing of intervening sequences: a new yeast mutant which fails to excise intervening sequences from precursor tRNAs. Cell 19:741–751
Huang B, Johansson MJ, Bystrom AS (2005) An early step in wobble uridine tRNA modification requires the Elongator complex. RNA 11:424–436
Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK (2003) Global analysis of protein localization in budding yeast. Nature 425:686–691
Hurt DJ, Wang SS, Lin YH, Hopper AK (1987) Cloning and characterization of LOS1, a Saccharomyces cerevisiae gene that affects tRNA splicing. Mol Cell Biol 7:1208–1216
Jackman JE, Alfonzo JD (2013) Transfer RNA modifications: nature's combinatorial chemistry playground. Wiley Interdisc Rev RNA 4:35–48
Jiang HQ, Motorin Y, Jin YX, Grosjean H (1997) Pleiotropic effects of intron removal on base modification pattern of yeast tRNAPhe: an in vitro study. Nucleic Acids Res 25:2694–2701
Johansson MJ, Bystrom AS (2002) Dual function of the tRNA(m(5)U54)methyltransferase in tRNA maturation. RNA 8:324–335
Johnson PF, Abelson J (1983) The yeast tRNATyr gene intron is essential for correct modification of its tRNA product. Nature 302:681–687
Jones TE, Brown CL, Geslain R, Alexander RW, Ribas de Pouplana L (2008) An operational RNA code for faithful assignment of AUG triplets to methionine. Mol Cell 29:401–407
Jones TE, Ribas de Pouplana L, Alexander RW (2013) Evidence for late resolution of the aux codon box in evolution. J Biol Chem 288:19625–19632
Joshi AK, Cherayil JD (1987) Stabilisation of some of the protein synthesis components in the thermophilic fungus, Humicola lanuginosa. J Biosci 11:193–202
Kalhor HR, Clarke S (2003) Novel methyltransferase for modified uridine residues at the wobble position of tRNA. Mol Cell Biol 23:9283–9292
Kamenski P, Kolesnikova O, Jubenot V, Entelis N, Krasheninnikov IA, Martin RP, Tarassov I (2007) Evidence for an adaptation mechanism of mitochondrial translation via tRNA import from the cytosol. Mol Cell 26:625–637
Kassavetis GA, Bartholomew B, Blanco JA, Johnson TE, Geiduschek EP (1991) Two essential components of the Saccharomyces cerevisiae transcription factor TFIIIB: transcription and DNA-binding properties. Proc Natl Acad Sci USA 88:7308–7312
Kassavetis GA, Kumar A, Ramirez E, Geiduschek EP (1998) Functional and structural organization of Brf, the TFIIB-related component of the RNA polymerase III transcription initiation complex. Mol Cell Biol 18:5587–5599
Konevega AL, Soboleva NG, Makhno VI, Semenkov YP, Wintermeyer W, Rodnina MV, Katunin VI (2004) Purine bases at position 37 of tRNA stabilize codon-anticodon interaction in the ribosomal A site by stacking and Mg^{2+}-dependent interactions. RNA 10:90–101
Kotelawala L, Grayhack EJ, Phizicky EM (2008) Identification of yeast tRNA Um(44) 2'-O-methyltransferase (Trm44) and demonstration of a Trm44 role in sustaining levels of specific tRNA(Ser) species. RNA 14:158–169
Kowalak JA, Dalluge JJ, McCloskey JA, Stetter KO (1994) The role of posttranscriptional modification in stabilization of transfer RNA from hyperthermophiles. Biochemistry, 33:7869–7876.
Lamichhane TN, Blewett NH, Crawford AK, Cherkasova VA, Iben JR, Begley TJ, Farabaugh PJ, Maraia RJ (2013) Lack of tRNA modification isopentenyl-A37 alters mRNA decoding and causes metabolic deficiencies in fission yeast. Mol Cell Biol 33:2918–2929
Lecointe F, Simos G, Sauer A, Hurt EC, Motorin Y, Grosjean H (1998) Characterization of yeast protein Deg1 as pseudouridine synthase (Pus3) catalyzing the formation of psi 38 and psi 39 in tRNA anticodon loop. J Biol Chem 273:1316–1323
Lee SJ, Jiko C, Yamashita E, Tsukihara T (2011) Selective nuclear export mechanism of small RNAs. Curr Opin Struct Biol 21:101–108

Lim VI (1995) Analysis of action of the wobble adenine on codon reading within the ribosome. J Mol Biol 252:277–282

Lin CA, Ellis SR, True HL (2010) The Sua5 protein is essential for normal translational regulation in yeast. Mol Cell Biol 30:354–363

Lipowsky G, Bischoff FR, Izaurralde E, Kutay U, Schafer S, Gross HJ, Beier H, Gorlich D (1999) Coordination of tRNA nuclear export with processing of tRNA. RNA 5:539–549

Macintosh GC, Bariola PA, Newbigin E, Green PJ (2001) Characterization of Rny1, the Saccharomyces cerevisiae member of the T2 RNase family of RNases: unexpected functions for ancient enzymes? Proc Natl Acad Sci USA 98:1018–1023

Machnicka MA, Milanowska K, Osman Oglou O, Purta E, Kurkowska M, Olchowik A, Januszewski W, Kalinowski S, Dunin-Horkawicz S, Rother KM, Helm M, Bujnicki JM, Grosjean H (2013) Modomics: a database of RNA modification pathways–2013 update. Nucleic Acids Res 41:D262-D267

Madsen CT, Mengel-Jorgensen J, Kirpekar F, Douthwaite S (2003) Identifying the methyltransferases for m(5)U747 and m(5)U1939 in 23S rRNA using MALDI mass spectrometry. Nucleic Acids Res 31:4738–4746

Malik S, Guermah M, Roeder RG (1998) A dynamic model for PC4 coactivator function in RNA polymerase II transcription. Proc Natl Acad Sci USA 95:2192–2197

Maraia RJ, Lamichhane TN (2011) 3' processing of eukaryotic precursor tRNAs. Wiley Interdisc Rev RNA 2:362–375

Mavrich TN, Ioshikhes IP, Venters BJ, Jiang C, Tomsho LP, Qi J, Schuster SC, Albert I, Pugh BF (2008) A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. Genome Res 18:1073–1083

McCloskey JA, Graham DE, Zhou S, Crain PF, Ibba M, Konisky J, Soll D, Olsen GJ (2001) Post-transcriptional modification in archaeal tRNAs: identities and phylogenetic relations of nucleotides from mesophilic and hyperthermophilic Methanococcales. Nucleic Acids Res 29:4699–4706

Melton DA, de Robertis EM, Cortese R (1980) Order and intracellular location of the events involved in the maturation of a spliced tRNA. Nature 284:143–148

Moqtaderi Z, Wang J, Raha D, White RJ, Snyder M, Weng Z, Struhl K (2010) Genomic binding profiles of functionally distinct RNA polymerase III transcription complexes in human cells. Nat Struct Mol Biol 17:635–640

Mori T, Ogasawara C, Inada T, Englert M, Beier H, Takezawa M, Endo T, Yoshihisa T (2010) Dual functions of yeast tRNA ligase in the unfolded protein response: unconventional cytoplasmic splicing of HAC1 pre-mRNA is not sufficient to release translational attenuation. Mol Biol Cell 21:3722–3734

Motorin Y, Helm M (2010) tRNA stabilization by modified nucleotides. Biochemistry 49:4934–4944

Moura GR, Paredes JA, Santos MA (2010) Development of the genetic code: insights from a fungal codon reassignment. FEBS Lett 584:334–341

Muramatsu T, Nishikawa K, Nemoto F, Kuchino Y, Nishimura S, Miyazawa T, Yokoyama S (1988) Codon and amino-acid specificities of a transfer RNA are both converted by a single post-transcriptional modification. Nature 336:179–181

Murphy FVT, Ramakrishnan V (2004) Structure of a purine-purine wobble base pair in the decoding center of the ribosome. Nat Struct Mol Biol 11:1251–1252

Murthi A, Shaheen HH, Huang HY, Preston MA, Lai TP, Phizicky EM, Hopper AK (2010) Regulation of tRNA bidirectional nuclear-cytoplasmic trafficking in Saccharomyces cerevisiae. Mol Biol Cell 21:639–649

Nishikura K, de Robertis EM (1981) RNA processing in microinjected Xenopus oocytes. Sequential addition of base modifications in the spliced transfer RNA. J Mol Biol 145:405–420

Noma K, Cam HP, Maraia RJ, Grewal SI (2006) A role for TFIIIC transcription factor complex in genome organization. Cell 125:859–872

Novoa EM, Pavon-Eternod M, Pan T, Ribas de Pouplana L (2012) A role for tRNA modifications in genome structure and codon usage. Cell 149:202–213

Novoa EM, Ribas de Pouplana L (2012) Speeding with control: codon usage, tRNAs, and ribosomes. Trends Genet 28:574–581

O'Connor JP, Peebles CL (1991) In vivo pre-tRNA processing in Saccharomyces cerevisiae. Mol Cell Biol 11:425–439

Oficjalska-Pham D, Harismendy O, Smagowicz WJ, Gonzalez de Peredo A, Boguta M, Sentenac A, Lefebvre O (2006) General repression of RNA polymerase III transcription is triggered by protein phosphatase type 2A-mediated dephosphorylation of Maf1. Mol Cell 22:623–632

Olivas WM, Muhlrad D, Parker R (1997) Analysis of the yeast genome: identification of new non-coding and small ORF-containing RNAs. Nucleic Acids Res 25:4619–4625

Orioli A, Pascali C, Pagano A, Teichmann M, Dieci G (2012) RNA polymerase III transcription control elements: themes and variations. Gene 493:185–194

Ozanick SG, Wang X, Costanzo M, Brost RL, Boone C, Anderson JT (2009) Rex1p deficiency leads to accumulation of precursor initiator tRNAMet and polyadenylation of substrate RNAs in Saccharomyces cerevisiae. Nucleic Acids Res 37:298–308

Parker R (2012) RNA degradation in Saccharomyces cerevisae. Genetics 191:671–702

Parrott AM, Mathews MB (2007) Novel rapidly evolving hominid RNAs bind nuclear factor 90 and display tissue-restricted distribution. Nucleic Acids Res 35:6249–6258

Parrott AM, Tsai M, Batchu P, Ryan K, Ozer HL, Tian B, Mathews MB (2011) The evolution and expression of the snaR family of small non-coding RNAs. Nucleic Acids Res 39:1485–1500

Patil A, Chan CT, Dyavaiah M, Rooney JP, Dedon PC, Begley TJ (2012) Translational infidelity-induced protein stress results from a deficiency in Trm9-catalyzed tRNA modifications. RNA Biol 9:990–1001

Phizicky EM, Hopper AK (2010) tRNA biology charges to the front. Genes Dev 24:1832–1860

Phizicky EM, Schwartz RC, Abelson J (1986) Saccharomyces cerevisiae tRNA ligase. Purification of the protein and isolation of the structural gene. J Biol Chem 261:2978–2986

Preston MA, Phizicky EM (2010) The requirement for the highly conserved G-1 residue of Saccharomyces cerevisiae tRNAHis can be circumvented by overexpression of tRNAHis and its synthetase. RNA 16:1068–1077

Purushothaman SK, Bujnicki JM, Grosjean H, Lapeyre B (2005) Trm11p and Trm112p are both required for the formation of 2-methylguanosine at position 10 in yeast tRNA. Mol Cell Biol 25:4359–4370

Putz J, Florentz C, Benseler F, Giege R (1994) A single methyl group prevents the mischarging of a tRNA. Nat Struct Biol 1:580–582

Qiu H, Hu C, Anderson J, Bjork GR, Sarkar S, Hopper AK, Hinnebusch AG (2000) Defects in tRNA processing and nuclear export induce GCN4 translation independently of phosphorylation of the alpha subunit of eukaryotic translation initiation factor 2. Mol Cell Biol 20:2505–2516

Rigby PW (1993) Three in one and one in three: it all depends on TBP. Cell 72:7–10

Roberts DN, Wilson B, Huff JT, Stewart AJ, Cairns BR (2006) Dephosphorylation and genome-wide association of Maf1 with Pol III-transcribed genes during repression. Mol Cell 22:633–644

Rocha R, Pereira PJ, Santos MA, Macedo-Ribeiro S (2011) Unveiling the structural basis for translational ambiguity tolerance in a human fungal pathogen. Proc Natl Acad Sci USA 108:14091–14096

Rose AM, Belford HG, Shen WC, Greer CL, Hopper AK, Martin NC (1995) Location of N2, N2-dimethylguanosine-specific tRNA methyltransferase. Biochimie 77:45–53

Rubio MA, Hopper AK (2011) Transfer RNA travels from the cytoplasm to organelles. Wiley Interdisc Rev RNA 2:802–817

Rudinger J, Florentz C, Giege R (1994) Histidylation by yeast HisRS of tRNA or tRNA-like structure relies on residues -1 and 73 but is dependent on the RNA context. Nucleic Acids Res 22:5031–5037

Samanta MP, Tongprasit W, Sethi H, Chin CS, Stolc V (2006) Global identification of noncoding RNAs in Saccharomyces cerevisiae by modulating an essential RNA processing pathway. Proc Natl Acad Sci USA 103:4192–4197

Sarkar S, Hopper AK (1998) tRNA nuclear export in Saccharomyces cerevisiae: in situ hybridization analysis. Mol Biol Cell 9:3041–3055

Shaheen HH, Hopper AK (2005) Retrograde movement of tRNAs from the cytoplasm to the nucleus in Saccharomyces cerevisiae. Proc Natl Acad Sci USA 102:11290–11295

Shibata S, Sasaki M, Miki T, Shimamoto A, Furuichi Y, Katahira J, Yoneda Y (2006) Exportin-5 orthologues are functionally divergent among species. Nucleic Acids Res 34:4711–4721

Silva RM, Paredes JA, Moura GR, Manadas B, Lima-Costa T, Rocha R, Miranda I, Gomes AC, Koerkamp MJ, Perrot M, Holstege FC, Boucherie H, Santos MA (2007) Critical roles for a genetic code alteration in the evolution of the genus Candida. EMBO J 26:4555–4565

Soma A, Ikeuchi Y, Kanemasa S, Kobayashi K, Ogasawara N, Ote T, Kato J, Watanabe K, Sekine Y, Suzuki T (2003) An RNA-modifying enzyme that governs both the codon and amino acid specificities of isoleucine tRNA. Mol Cell 12:689–698

Spinelli SL, Consaul SA, Phizicky EM (1997) A conditional lethal yeast phosphotransferase (tpt1) mutant accumulates tRNAs with a 2'-phosphate and an undermodified base at the splice junction. RNA 3:1388–1400

Srinivasan M, Mehta P, Yu Y, Prugar E, Koonin EV, Karzai AW, Sternglanz R (2011) The highly conserved KEOPS/EKC complex is essential for a universal tRNA modification, t6A. EMBO J 30:873–881

Stanford DR, Whitney ML, Hurto RL, Eisaman DM, Shen WC, Hopper AK (2004) Division of labor among the yeast Sol proteins implicated in tRNA nuclear export and carbohydrate metabolism. Genetics 168:117–127

Takano A, Endo T, Yoshihisa T (2005) tRNA actively shuttles between the nucleus and cytosol in yeast. Science 309:140–142

Tavenet A, Suleau A, Dubreuil G, Ferrari R, Ducrot C, Michaut M, Aude JC, Dieci G, Lefebvre O, Conesa C, Acker J (2009) Genome-wide location analysis reveals a role for Sub1 in RNA polymerase III transcription. Proc Natl Acad Sci USA 106:14265–14270

Thomas BC, Li X, Gegenheimer P (2000) Chloroplast ribonuclease P does not utilize the ribozyme-type pre-tRNA cleavage mechanism. RNA 6:545–553

Thompson DM, Lu C, Green PJ, Parker R (2008) tRNA cleavage is a conserved response to oxidative stress in eukaryotes. RNA 14:2095–2103

Thompson DM, Parker R (2009) Stressing out over tRNA cleavage. Cell 138:215–219

Torres AG, Batlle E, Ribas de Pouplana L (2014) Role of tRNA modifications in human diseases. Trends Mol Med. doi:10.1016/j.molmed.2014.01.008

Towns WL, Begley TJ (2012) Transfer RNA methytransferases and their corresponding modifications in budding yeast and humans: activities, predications, and potential roles in human health. DNA Cell Biol 31:434–454

Trotta CR, Miao F, Arn EA, Stevens SW, Ho CK, Rauhut R, Abelson JN (1997) The yeast tRNA splicing endonuclease: a tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases. Cell 89:849–858

Trotta CR, Paushkin SV, Patel M, Li H, Peltz SW (2006) Cleavage of pre-tRNAs by the splicing endonuclease requires a composite active site. Nature 441:375–377

Urbonavicius J, Auxilien S, Walbott H, Trachana K, Golinelli-Pimpaneau B, Brochier-Armanet C, Grosjean H (2008) Acquisition of a bacterial RumA-type tRNA(uracil-54, C5)-methyltransferase by Archaea through an ancient horizontal gene transfer. Mol Microbiol 67:323–335

Urbonavicius J, Qian Q, Durand JM, Hagervall TG, Bjork GR (2001) Improvement of reading frame maintenance is a common function for several tRNA modifications. EMBO J 20:4863–4873

Urbonavicius J, Stahl G, Durand JM, Ben Salem SN, Qian Q, Farabaugh PJ, Bjork GR (2003) Transfer RNA modifications that alter +1 frameshifting in general fail to affect −1 frameshifting. RNA 9:760–768

van Hoof A, Staples RR, Baker RE, Parker R (2000) Function of the ski4p (Csl4p) and Ski7p proteins in 3′-to-5′ degradation of mRNA. Mol Cell Biol 20:8230–8243
Vannini A, Cramer P (2012) Conservation between the RNA polymerase I, II, and III transcription initiation machineries. Mol Cell 45:439–446
Waas WF, Druzina Z, Hanan M, Schimmel P (2007) Role of a tRNA base modification and its precursors in frameshifting in eukaryotes. J Biol Chem 282:26026–26034
Walker SC, Engelke DR (2008) A protein-only RNase P in human mitochondria. Cell 135:412–414
Wallrath LL, Geyer PK (2006) TFIIIC boxes in the genome. Cell 125:829–831
Wang Z, Roeder RG (1998) DNA topoisomerase I and PC4 can interact with human TFIIIC to promote both accurate termination and transcription reinitiation by RNA polymerase III. Mol Cell 1:749–757
Werten S, Moras D (2006) A global transcription cofactor bound to juxtaposed strands of unwound DNA. Nat Struct Mol Biol 13:181–182
Whipple JM, Lane EA, Chernyakov I, D'Silva S, Phizicky EM (2011) The yeast rapid tRNA decay pathway primarily monitors the structural integrity of the acceptor and T-stems of mature tRNA. Genes Dev 25:1173–1184
White RJ (2011) Transcription by RNA polymerase III: more complex than we thought. Nat Rev Genet 12:459–463
White RJ, Rigby PW, Jackson SP (1992) The TATA-binding protein is a general transcription factor for RNA polymerase III. J Cell Sci Suppl 16:1–7
Whitney ML, Hurto RL, Shaheen HH, Hopper AK (2007) Rapid and reversible nuclear accumulation of cytoplasmic tRNA in response to nutrient availability. Mol Biol Cell 18:2678–2686
Willis IM (1993) RNA polymerase III. Genes, factors and transcriptional specificity. Eur J Biochem 212:1–11
Willis IM (2002) A universal nomenclature for subunits of the RNA polymerase III transcription initiation factor TFIIIB. Genes Dev 16:1337–1338
Wolf J, Gerber AP, Keller W (2002) tadA, an essential tRNA-specific adenosine deaminase from Escherichia coli. EMBO J 21:3841–3851
Wolfe CL, Hopper AK, Martin NC (1996) Mechanisms leading to and the consequences of altering the normal distribution of ATP(CTP):tRNA nucleotidyltransferase in yeast. J Biol Chem 271:4679–4686
Wolfe CL, Lou YC, Hopper AK, Martin NC (1994) Interplay of heterogeneous transcriptional start sites and translational selection of AUGs dictate the production of mitochondrial and cytosolic/nuclear tRNA nucleotidyltransferase from the same gene in yeast. J Biol Chem 269:13361–13366
Yang L, Altman S (2007) A noncoding RNA in Saccharomyces cerevisiae is an RNase P substrate. RNA 13:682–690
Yoo CJ, Wolin SL (1997) The yeast La protein is required for the 3' endonucleolytic cleavage that matures tRNA precursors. Cell 89:393–402
Yoshihisa T, Ohshima C, Yunoki-Esaki K, Endo T (2007) Cytoplasmic splicing of tRNA in Saccharomyces cerevisiae. Genes Cells 12:285–297
Yoshihisa T, Yunoki-Esaki K, Ohshima C, Tanaka N, Endo T (2003) Possibility of cytoplasmic pre-tRNA splicing: the yeast tRNA splicing endonuclease mainly localizes on the mitochondria. Mol Biol Cell 14:3266–3279
Zhu L, Deutscher MP (1987) tRNA nucleotidyltransferase is not essential for Escherichia coli viability. EMBO J 6:2473–2477

Chapter 11
Small RNA-Mediated Gene Silencing in *Neurospora*

Yunkun Dang, Zhenyu Zhang and Yi Liu

Abstract *Neurospora crassa* is a filamentous fungus that serves as a eukaryotic model system for the study of many important biological processes. One of the earliest known examples of RNA silencing was discovered in *Neurospora*, and it has become an important experimental system for analyses of the mechanisms of RNA interference and small RNA biogenesis. *Neurospora* possesses two small RNA-based genome defense mechanisms, quelling and meiotic silencing of unpaired DNA. The known biogenesis pathways of small RNAs, including microRNA and DNA-damage-induced small RNAs, are surprisingly diverse. The studies of RNAi pathways and small RNA production in *Neurospora* have provided important insights into our understanding of small RNA-mediated gene silencing mechanisms in eukaryotic organisms.

Contents

Y. Dang · Z. Zhang · Y. Liu (✉)
Department of Physiology, ND13.214A, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75390–9040, USA
e-mail: Yi.Liu@UTsouthwestern.edu

A. Sesma and T. von der Haar (eds.), *Fungal RNA Biology*,
DOI: 10.1007/978-3-319-05687-6_11,

The Discovery of RNA Interference

RNA interference (RNAi) refers to a highly conserved gene silencing process in eukaryotes. In this process, small RNA molecules ranging from 20 to 30 nucleotides (nt) in length are critical for regulation of gene expression because of their complementarity to mRNAs. RNAi is one of the most important biological discoveries in the past 20 years. Small interfering RNAs participate in numerous biological processes ranging from development and cancer to host defense, and they also serve as invaluable research tools to study gene function. RNAi-based technologies are extensively used in pharmaceutical target validation and are being evaluated in late-stage clinical trials.

In hindsight, RNAi-related phenomena were probably first reported in the mid-1980s, when the inhibitory effects of exogenously introduced antisense RNAs to target genes expressed in various organisms were observed (Crowley et al. 1985; Kim and Wold 1985; Ecker and Davis 1986). The RNAi pathway was first elucidated in studies in plants and fungi in the 1990s. In a study of the flower color of petunias, Napoli et al. introduced extra copies of the chalcone synthase gene, a key enzyme required for flower pigmentation to confer pink or violet color, into petunias. To their surprise, rather than getting flowers with darker pigmentation, the repetitive transgene resulted in light color or even white flowers, indicating that the transgene had silenced both the ectopic and the endogenous chalcone synthase genes (1990). Further investigations of the phenomenon, termed "co-suppression," indicated that the transgene resulted in post-transcriptional gene silencing via an increased rate of mRNA degradation (Van Blokland et al. 1994).

In 1992, a similar phenomenon was discovered by Macino and colleagues in the filamentous fungus *Neurospora crassa*. When Romano and Macino transformed the *albino-1(al-1)* gene, which encodes a key enzyme for carotenoid biosynthesis that confers orange color into *Neurospora*, they unexpectedly found many pale yellow and white transformants, indicating the silencing of the ectopic and endogenous *al-1* gene (Romano and Macino 1992). They termed this phenomenon "quelling," but the mechanistic relation to the co-suppression in plants was not immediately clear at the time. By studying the mechanisms of quelling and by obtaining *Neurospora* mutants deficient in quelling, Macino, Cogoni, and their colleagues demonstrated that quelling is a post-transcriptional gene silencing mechanism mediated by RNA (see below).

In 1997, virus-induced gene silencing was discovered in plants (Covey et al. 1997; Ratcliff et al. 1997). Plants that carried only short, non-coding regions of viral RNA sequences were protected from viral infection. In addition, when short sequences of plant genes were introduced into viruses, the expression of the homologous gene was suppressed in a virus-infected plant. These results suggested the potential role of RNAi as a host-defense mechanism against viral infection. This phenomenon, together with other similar ones, including quelling in *Neurospora*, was called "post-transcriptional gene silencing."

The landmark breakthrough for the understanding of RNAi came in 1998 when Craig Mello and Andrew Fire reported that double-stranded RNA (dsRNA) triggers a potent gene silencing effect in *Caenorhabditis elegans*. They found that neither mRNA nor antisense RNA had a significant effect on gene expression, but dsRNA robustly silenced the targeted gene (Fire et al. 1998). They termed this phenomenon as "RNA interference." This discovery won Fire and Mello the Nobel Prize for Physiology and Medicine in 2006 and was particularly notable because they revealed the trigger of post-transcriptional gene silencing and established the basis of RNAi. Soon after their report, Macino and colleagues cloned the first gene required for quelling, *qde-1*, which has sequence homology to cellular RNA-dependent RNA polymerase, an enzyme that synthesizes dsRNA from single-stranded RNA (ssRNA) (Cogoni and Macino 1999a). This seminal discovery in *Neurospora* established dsRNA as the endogenous trigger for RNAi in eukaryotes and revealed the shared mechanistic link among different gene silencing phenomena in different organisms.

The studies of RNAi processes in different model systems established it as a highly conserved mechanism across all kingdoms of eukaryotes. In general, small ($\sim$20–30 nucleotide) non-coding RNAs, which are mostly products of dsRNA, recognize and silence mRNAs harboring homologous sequences via RNA cleavage, translational repression, or induction of heterochromatin formation (Hamilton and Baulcombe 1999; Zamore et al. 2000; Bernstein et al. 2001; Elbashir et al. 2001). Three major classes of small RNAs have been identified in eukaryotes: short interfering RNAs (siRNAs), microRNAs (miRNAs), and Piwi-interacting RNAs (piRNAs). siRNA and miRNA have many features in common, except that siRNAs are made from dsRNA precursors, whereas miRNAs are produced from transcripts harboring a hairpin structure. siRNAs and miRNAs are 21–25 base-pair duplexes generated from double-stranded precursors by Dicer, a ribonuclease III (RNaseIII) enzyme. The two strands of the miRNA and siRNA duplexes are known as the guide strand (the strand with complementarity to mRNA) and the passenger strand. These RNA duplexes are loaded onto the RNA-induced silencing complex (RISC) or a nuclear form of RISC called the RNA-induced transcriptional silencing complex (RITS) (Hammond et al. 2001; Verdel et al. 2004). RISC and RITS are effector complexes that are targeted at homologous RNA sequences by base-pairing interactions involving the guide strand of the small RNA duplex. The core component of each complex is a conserved PAZ and PIWI domain-containing protein called Argonaute, which binds to the guide RNA and cleaves the target RNA using its RNaseH-like PIWI domain (Wang et al. 2008a, b; Nakanishi et al. 2012). piRNAs, on the other hand, are very different from miRNAs and siRNAs. piRNAs are only found in animal cells and are distinct in size (26–31 nt). piRNAs bind to PIWI domain-containing proteins and cause epigenetic and post-transcriptional gene silencing of retrotransposons and other genetic elements in germline cells (reviewed in Luteijn and Ketting 2013).

Neurospora

The filamentous fungus *N. crassa* belongs to the phylum *Ascomycota*; these organisms live mainly in tropical and subtropical regions. *Neurospora* is easy to grow and has a haploid life cycle that makes genetic analysis simple since recessive traits appear in the offspring. Due to these advantages, *Neurospora* has been used as a genetic model since the 1920s. In the 1940s, Beadle and Tatum used *N. crassa* to study metabolic pathways, which led them to propose the "one gene, one enzyme" theory, winning them the Nobel Prize in 1958. Since then, *Neurospora* has been used as a powerful model system to study many biological processes including the circadian clock, photobiology, epigenetics, and RNAi.

The life cycle of *N. crassa* can be divided into asexual and sexual cycles (Fig. 11.1). *N. crassa* has two mating types (*A* and *a*) that are morphologically indistinguishable from each other. The vegetative phase is initiated when a spore, either a conidium produced from an asexual cycle or an ascospore produced during a sexual cycle, geminates into multicellular filaments (hyphae) and further develops into a mycelium, consisting of a mass of branching hyphae. The germination of ascospores requires heat, such as forest fire in the wild, whereas conidia germinate spontaneously. After the mycelium is well established, aerial hyphae (conidiophores) develop, leading to the production of the abundant orange conidia. The conidia, which contain one to several nuclei each, can either establish new vegetative cultures or fertilize strains of the opposite mating type.

The sexual cycle of *Neuropsora* is initiated if nutrients are limiting. The female strains, which can be derived from both mating types, produce fruiting bodies called protoperithecia from which specialized hypha (trichogyne) protrude. When a trichogyne contacts tissue of the opposite mating type (male), it acquires the male nucleus and transports it back to the protoperithecium which develops into perithecium. The fertilized heterokaryotic cells proliferate in the perithecium and form ascogenous (heterokaryotic) and maternal (homokaryotic) tissues. The former first develop into hook-shaped cells called croziers that then develop into three cells. The middle cell, also known as ascus mother cell, carries both male and female nuclei and will subsequently undergo karyogamy (nuclear fusion), meiosis, post-meiotic mitosis, and eventually form an ascus cell that contains eight haploid spores (ascospores) arrayed in an order that reflects their lineage (Raju 1980, 1992). The diploid phase of the life cycle is brief ($\sim$24 h) and limited to a single developing cell. Ascospores are ejected from the beak of the perithecium and can germinate after exposure to high temperature to produce vegetative mycelia, completing the sexual cycle.

The vegetative stage (asexual cycle) of *Neurospora* accounts for most of its life cycle. Hyphae, the vegetative tissue of filamentous fungi, are typically divided into cells by internal cross-walls called septa (Harris 2001). Interestingly, of the $\sim$40-megabase *N. crassa* genome, only $\sim$9 % of the genome is repetitive DNA (Galagan et al. 2003). Apart from a tandem array of approximately 200 copies of the $\sim$10-kb ribosomal DNA (rDNA) unit encoding the three large ribosomal

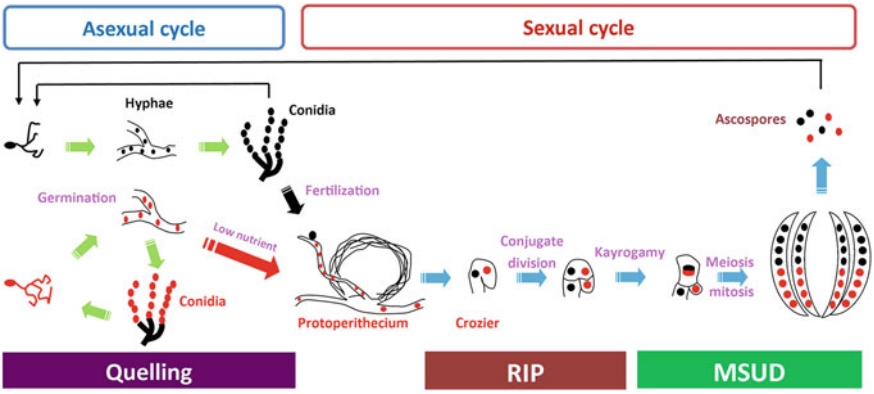


Fig. 11.1 Life cycle of *Neurospora*. The *asexual cycle* depicts the germination of conidia or ascospores, the formation of hyphae, and production of conidia. The *sexual cycle* depicts the maturation of ascus within a perithecium. The different mating types are indicated with *black* (A) and *red* (a) dots. The *boxes* below the life cycle cartoon indicate the timing of occurrence of three homology-based defense systems, quelling, RIP, and MSUD

RNAs, most of the repetitive DNA consists of inactivated and heavily mutated transposable elements. These results reflect the fact that *N. crassa* has at least three potent homology-based genomic defense systems (Fig. 11.1). The first, quelling, occurs during the vegetative stage. The second mechanism, meiotic silencing of unpaired DNA (MSUD), occurs during the sexual stage, specifically after the karyogamy when two nuclei of different mating types are fused together (Shiu et al. 2001). The third defense mechanism is known as repeat-induced point mutation (RIP); RIP is initiated during sexual stage (before karyogamy) and maintains its silencing effect across the whole life cycle by generating point mutations in repetitive sequences (Galagan and Selker 2004). The common feature of these three genomic defense processes in *Neurospora* is that they all rely on the recognition of homologous and repetitive DNA sequence. Quelling and RIP are induced by repetitive sequences, whereas MSUD is initiated when there are unpaired homologous chromosomes during sexual phase. We will not cover RIP in this review, since RNAi is known not to be involved in this process (Freitag et al. 2004). *Neurospora* also produces several types of small RNAs through different biogenesis pathways: DNA damage-induced *QDE-2* associated siRNA (qiRNA), miRNA-like small RNA (milRNA) and Dicer-independent siRNA (disiRNA). The diverse small RNA biogenesis pathways in *Neurospora*, along with the importance of RNA-based genomic defense mechanisms and its genetic tractability make *Neurospora* an excellent model system for the study of RNAi and small RNA biogenesis pathways.

Quelling

The discovery of quelling was a landmark event in RNAi research. Quelling was discovered by transformation of *Neurospora* with carotenoid biosynthesis genes, *al-1* and *al-3*, which resulted in silencing of endogenous *al-1* and *al-3* genes, as indicated by the albino/pale yellow phenotype and reduction in *al* mRNA levels in around 30 % of the transformants (Romano and Macino 1992). This silencing phenotype spontaneously and progressively reverts to wild type or intermediate phenotypes after prolonged culture time (Romano and Macino 1992; Fulci and Macino 2007). The exogenous transgene needs to be longer than ~132 nt in order to have a silencing effect. The initiation of quelling is independent of promoter sequences or other specific sequences, suggesting that it is likely to be triggered by some aberrant DNA structure. Furthermore, there is a strong correlation between the copy number of transgenes and the silencing efficiency. In fact, progressive phenotypic reversion of quelled strains correlates with the reduction of the number of the ectopic integrated sequences (Romano and Macino 1992; Cogoni et al. 1994, 1996), suggesting that repetitive transgenes result in gene silencing.

Quelling was proposed to be mediated by RNA based on several observations. First, mutations in carotenoid genes are typically recessive; however, most of the *al-1* quelled strains were heterokaryons and were dominant over wild-type strains, indicating that a diffusible, *trans*-acting molecule was involved in quelling (Cogoni et al. 1996). Second, a sense RNA derived from the exogenous promoter-less *al-1* transgene was present in quelled strains but absent in the reverted strains, suggesting that transcription of the transgenes was involved in the quelling pathway (Cogoni et al. 1996). In addition, quelling acted at the post-transcriptional level because the amount of precursor mRNA was about the same in both quelled and non-quelled strains, whereas the amount of mature mRNA was reduced in quelled strains (Cogoni et al. 1996). These findings together led to the hypothesis that the production of aberrant RNA (aRNA) in the presence of multiple copies of a transgene causes post-transcriptional gene silencing in a *trans*-acting manner. These are one of the earliest studies suggesting that an aRNA transcript was involved in a gene silencing mechanism.

Several components of the quelling pathway have been identified by either forward- or reverse-genetic approaches (Cogoni and Macino 1997; Catalanotto et al. 2004; Dang et al. 2011). Forward genetic screens led to the isolation of three complementation groups of quelling-defective (*qde*) mutants, called *qde-1*, *qde-2*, and *qde-3*. Studies of these three genes, together with findings from other organisms, established the framework of the quelling pathway as an RNAi-related mechanism that is well conserved across eukaryotes (Cogoni and Macino 1997). *QDE-1*, the first RNAi component identified, is homologous to RNA-dependent RNA polymerase (RdRP) (Cogoni and Macino 1999a). The requirement for an RdRP in quelling demonstrates that dsRNA is a necessary intermediate in this gene silencing pathway.

QDE-3 belongs to the *RecQ* helicase family. *RecQ* helicases are involved in homologous recombination, DNA replication, and DNA repair (Cogoni and Macino 1999b). *QDE-3* and another *RecQ* helicase homolog, *RecQ2*, are necessary for DNA repair in *Neurospora* (Pickford et al. 2003; Kato et al. 2004). It was reasoned that quelling is not due to specific sequences since a large variety of DNA fragments can induce quelling without promoters. Thus, it is likely that aberrant chromatin structures adopted by repetitive transgenes facilitate the production of siRNA. The involvement of *QDE-3* in quelling suggests that repetitive transgenes can be recognized and processed by *QDE-3* to produce an aberrant DNA structure that is recognized to produce siRNA. Interestingly, *OsRecQ1*, a *RecQ* helicase homolog in rice, is required for inverted-repeat-induced RNA silencing (Chen et al. 2008), and *rRecQ-1*, the rat homolog of *QDE-3*, is associated with piRNA-binding complex (Lau et al. 2006).

Because inverted repeat-containing transgenes can bypass *QDE-1* and *QDE-3* to produce dsRNA (Li et al. 2010a), *QDE-1* and *QDE-3* function upstream in the quelling pathway and are likely to be involved in the production of aberrant RNA. A question that arises here is how *QDE-1* and/or *QDE-3* produce RNA from a DNA template. In plants, it has been demonstrated that RNA polymerases IV and V (Pol IV and Pol V, respectively) are required for the production of non-coding RNA, which is the precursor of endogenous siRNA (Wierzbicki et al. 2008). In the fission yeast *Schizosaccharomyces pombe*, Pol II synthesizes the centromeric siRNA precursor (Djupedal et al. 2005; Schramke et al. 2005). The crystal structure of *QDE-1* revealed that its catalytic core is similar to that of eukaryotic DNA-dependent RNA polymerase (DdRP) (Salgado et al. 2006). Our lab demonstrated that *QDE-1* has both DdRP and RdRP functions, utilizing single-stranded DNA (ssDNA) and ssRNA as templates to produce DNA/RNA hybrids and dsRNAs, respectively (Lee et al. 2010b). Interestingly, the DdRP activity of *QDE-1* is much higher than its RdRP activity in vitro. This is consistent with the observation that *QDE-1* is required for aRNA production and is not involved in the amplification and production of secondary small RNAs, unlike some RdRPs in other organisms (Lee et al. 2009). *QDE-1* has been shown to associate with *RPA-1*, the largest subunit of replication protein A (*RPA*), which binds ssDNA and is involved in DNA replication, recombination, and repair (Nolan et al. 2008; Lee et al. 2010b). *RPA* strongly promotes dsRNA production by *QDE-1* from ssDNA by preventing the formation of a DNA/RNA hybrid. These results suggest that *RPA* has dual roles in the production of aRNA: it recruits *QDE-1* to ssDNA and blocks the formation of DNA/RNA hybrids.

After the production of a dsRNA precursor, two RNaseIII domain-containing proteins, Dicer-like-1 (*Dcl-1*) and *Dcl-2*, process this precursor into small duplex RNAs of about 25 base pairs in an ATP-dependent manner (Catalanotto et al. 2004). The elimination of both *dicer* genes completely abolishes quelling and the processing of dsRNA into the siRNA form. However, single mutants have quelling frequencies comparable to that of the wild-type strain, indicating that these two Dicers are functionally redundant. In vitro assays suggest that *Dcl-2* is the major

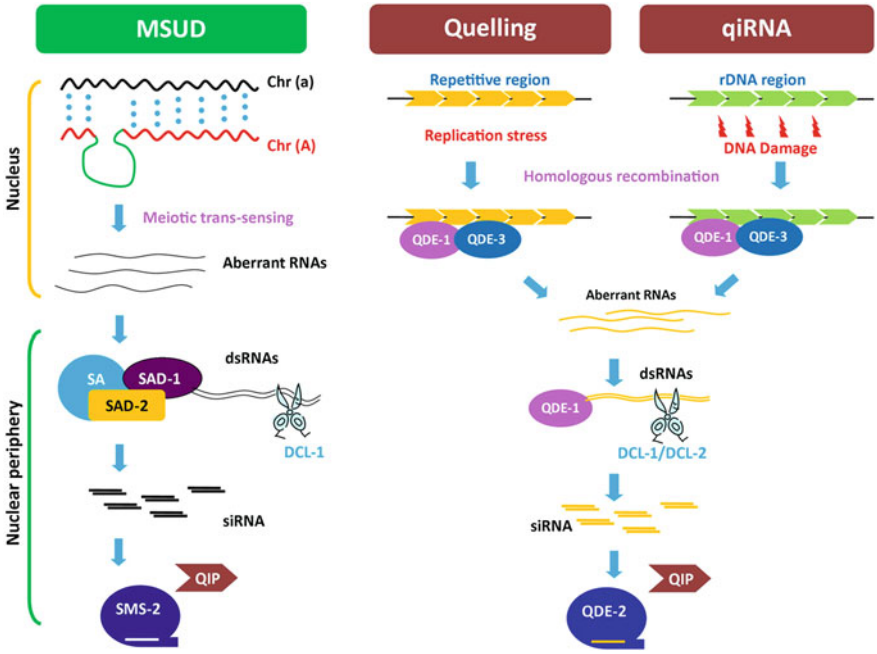


Fig. 11.2 The models of siRNA-related biogenesis pathways in *Neurospora*. During vegetative growth, when repetitive sequences (quelling, *middle panel*) and ribosomal DNA regions (qiRNA, *right panel*) undergo replication stress and DNA damage, respectively, homologous recombination is initiated and aRNAs are produced by *QDE-1* and *QDE-3*. The aRNAs are then converted into dsRNAs with *QDE-1*, cleaved by *Dcl-1* or *Dcl-2* into siRNA, which are loaded onto *QDE-2*. QIP removes the passenger strand, resulting in the mature RISC complex. During meiosis (*left*), unpaired DNA is sensed by comparing the homologous chromosomes of two different mating type strains, resulting in the production of aRNAs. These aRNAs are further converted into dsRNAs by *Sad-1*, *-2* and *-3*, processed by *Dcl-1* and then loaded onto *Sms-2*. QIP are presumably required for the maturation of RISC complex by removing the passenger strand, as in quelling

dsRNA processing enzyme because the *Dcl-2* mutant but not *Dcl-1* mutant has reduced Dicer activity (Catalanotto et al. 2004).

QDE2, an Argonaute homolog in *Neurospora*, associates with a siRNA duplex to form an inactive RISC (Catalanotto et al. 2000, 2002). Consistent with this, dsRNA and siRNA production depends on functional *QDE-1* and *QDE-3* but not *QDE-2*, indicating that *QDE-2* functions in a downstream step of the gene silencing pathway (Catalanotto et al. 2002) (Fig. 11.2). To activate RISC and execute gene silencing, the passenger strand of the siRNA duplex must be removed. We showed that *QDE-2* and its slicer activity are required for production of the single-stranded form of the siRNA and for gene silencing in vivo (Maiti et al. 2007). This provided the first in vivo evidence that Argonaute was involved in generating single-stranded siRNA and RISC activation. However, *QDE-2* alone is not sufficient to remove the passenger strand. Biochemical purification of *QDE-2* led to the identification of a *QDE-2* interacting protein (QIP), which is an

exonuclease (Maiti et al. 2007). Further in vitro experiments indicated that QIP cleaves and removes the nicked passenger strand from siRNA duplex in a *QDE-2* dependent manner. Thus, QIP plays a critical role in dsRNA-induced gene silencing and was one of the first identified eukaryotic exonucleases required for efficient RNAi. A *Drosophila* ribonuclease, C3PO (component 3 promoter of RISC), also promotes RISC activation by removing siRNA passenger strand cleavage products (Liu et al. 2009). These results indicate that cleavage and removal of passenger strand from the siRNA duplex is an essential step in dsRNA-induced gene silencing in eukaryotes.

Quelling is triggered in the presence of repetitive transgenes in *Neurospora* genome, which indicate that it is a mechanism that functions in genome defense by suppressing transposon activity. Due to the RIP process during the sexual cycle (Galagan and Selker 2004), most of the repetitive genes have been mutated in *Neurospora*, and most of the *Neurospora* strains examined have no functional transposons. One functional LINE-like transposon, Tad, from an African strain was identified (Kinsey 1990). Nolan et al. introduced this Tad transposon into the *Neurospora* laboratory strain and showed that repression of its activity requires *QDE-2* and Dicer, but not *QDE-1* or *QDE-3* (Nolan et al. 2005). These results suggest that transposition of Tad may generate inverted repeats that form dsRNA without the requirement of *QDE-1* and *QDE-3*.

In summary, to initiate quelling, repetitive transgenes in quelled strains forms aberrant DNA structures. *QDE-3* and *RPA* are involved in this process. The aberrant DNA structures are recognized and transcribed by *QDE-1* to produce aberrant RNA and then dsRNA. These long dsRNA precursors are processed into siRNA duplexes by Dicer proteins, and siRNAs are loaded onto the RISC. *QDE-2* cleaves and removes the passenger strand with the help of QIP to form an active RISC associated with a single-stranded siRNA, resulting in gene silencing (Fig. 11.2).

qiRNA Biogenesis Pathway

In addition to quelling-induced siRNA, we identified another class of endogenous siRNAs synthesized during vegetative stage after DNA damage agent treatments (Lee et al. 2009). Because this type of small RNA is associated with *QDE-2*, they were called *QDE-2*-interacting small RNAs or qiRNAs (Lee et al. 2009). Statistical and biochemical analysis of qiRNAs revealed that these small RNAs have a strong preference of 5′ uridine and 3′ adenine. Deep sequencing analysis showed that most of the qiRNAs (~90 %) originate from the ribosomal DNA (rDNA) cluster. It should be noted here that rDNA remains the only highly repetitive region in the *Neurospora* genome due to the existence of repeat-target gene silencing/mutation mechanisms such as MSUD and RIP (Aramayo and Metzenberg 1996; Selker 2002).

Interestingly, qiRNAs are produced not only from the transcribed region of rDNA but also from the untranscribed intergenic spacer regions, indicating that

qiRNAs originate from aRNA precursors. Northern blot analyses revealed that these aRNA precursors range in size from approximately 500 bp–2 kbp. The production of qiRNA does not depend on Pol I, Pol II, or Pol III. Genetic analysis showed that qiRNA biogenesis requires *QDE-1*, *QDE-3*, and Dicers, similar to the production of siRNAs involved in the quelling pathway (Lee et al. 2009). These results suggest that qiRNA is specifically made by the RNAi machinery. In the *dicer-1*/*dicer-2* double knock-out strain, aRNA accumulates, indicating that Dicers function in the processing of aRNA into these small dsRNA (Lee et al. 2009). *qde-1* and *qde-3* mutants abolished the production of aRNA, indicating that the *RecQ* helicase and the RdRP/DdRP *QDE-1* are required for the biogenesis of DNA damage-induced aRNA. This result is consistent with the notion that *QDE-1* is a DdRP that generates single-strand aRNA from genomic loci (Salgado et al. 2006; Lee et al. 2009, 2010b).

qiRNA production can be induced by a large variety of DNA damage agents, including hydroxyurea, camptothecin, histidine, methyl methanesulfonate, and ethyl methanesulfonate, which cause DNA damage through different mechanisms (Lee et al. 2009; Zhang et al. 2013). Moreover, mutant strains that are deficient in DNA damage repair or checkpoint pathways showed elevated levels of qiRNA without DNA damage agent treatment (Lee et al. 2009). Because the external or internal cues that induce qiRNA production ultimately cause DNA double-strand breaks, these results suggest that double-stranded breaks in the DNA are the trigger for qiRNA induction. Since the discovery of qiRNA, recent studies in plants, flies, and mammals (Francia et al. 2012; Michalik et al. 2012; Wei et al. 2012; Sharma and Misteli 2013) suggest that DNA damage is a common trigger for siRNA production in eukaryotes; however, how DNA damage is sensed and how it promotes the production of aRNA remain largely unclear.

Convergence of Quelling and qiRNA Pathways

The initiation of quelling does not require specific DNA sequences, indicating that the repetitive nature of sequences plays a key role in triggering the aRNA production (Romano and Macino 1992; Li et al. 2010b; Chang et al. 2012). Both the quelling and the qiRNA pathways generate aRNA from repetitive sequences with the involvement of *QDE-1*, *QDE-3*, and RPA (Nolan et al. 2008; Lee et al. 2010b; Zhang et al. 2013). To understand how repetitive sequences are recognized to produce siRNA and to identify novel quelling components, we carried out a genetic screen for new quelling/qiRNA pathway mutants using the *Neurospora* knock-out library (Colot et al. 2006; Zhang et al. 2013). We found that DNA damage-induced qiRNA production was abolished in mutants deficient in key homologous recombination (HR) components including *Rad51*, *Rad52*, and *Rad54* (Zhang et al. 2013). Mutants deficient in other DNA damage repair and checkpoint pathways, however, were not deficient in qiRNA biogenesis. Quelling assays revealed that HR is also required for the quelling pathway (Zhang et al. 2013). These results suggest that HR serves as the mechanism to distinguish repetitive

sequences from the rest of the genome because repetitive sequences will have more frequent homologous recombination. These results further suggest that the upstream biogenesis of qiRNA and the small RNAs (sRNAs) involved in quelling is mechanistically similar.

Quelling and the qiRNA pathway share several other features in common, although quelling can occur under normal growth conditions without DNA damage agent treatment. Both pathways require the same set of genetic components to generate siRNAs, and the siRNAs involved in quelling and in the qiRNA pathway originate from repetitive sequences (Lee et al. 2009; Zhang et al. 2013). More importantly, we found that quelling-induced siRNA can also be enhanced by DNA damage agent treatment (Zhang et al. 2013), indicating that quelling-induced siRNAs are also a result from DNA damage. Repetitive sequences are known to be a major source of genome instability due to recombination (Bzymek and Lovett 2001; Vader et al. 2011) and repeat-induced gene silencing has been reported in many organisms including yeast, plants, and mammals (Napoli et al. 1990; Romano and Macino 1992; Hsieh and Fire 2000; Chang et al. 2012). The study of siRNA biogenesis in *Neurospora* thus will provide important insights into repetitive DNA trigged gene silencing in other eukaryotes.

The major difference between quelling and the qiRNA pathway is that qiRNA production is triggered by DNA damage, whereas quelling takes place under normal growth conditions (Romano and Macino 1992; Lee et al. 2009). The repetitive rDNA cluster does not give rise to the production of qiRNA under normal growth conditions, indicating that the rDNA region is somehow protected from HR in the absence of DNA damage. Previous studies have shown that several mechanisms, including rDNA transcriptional silencing and regulation of rDNA replication, prevent rDNA hyper-recombination and maintain rDNA copy numbers (Calzada et al. 2005; Huang et al. 2006). The replication fork protection complex has been shown to be important for maintenance of rDNA stability by preventing the replication fork collapse (Krings and Bastia 2004; Mohanty et al. 2006). Interestingly, mutants deficient in replication fork protection complex produce qiRNA in the absence of DNA damage agent treatment (Zhang et al. 2013). These results suggest that the difference between quelling and qiRNA is that rDNA cluster is normally protected from HR, whereas the repetitive transgenic loci are not.

Meiotic Silencing of Unpaired DNA

The MSUD pathway silences gene expression during the meiotic stage through an RNAi-based mechanism (Aramayo et al. 1996; Shiu et al. 2001). MSUD acts in the prophase of first meiosis, when there is a region of unpaired DNA between the two homologous parental chromosomes (i.e., DNA exists in one parental chromosome but not in its pairing partner or in heterologous DNA on parental chromosomes); this can also be resulted from the presence of repetitive DNA in one parental strain during the mating process. MSUD was originally called meiotic transvection or

meiotic *trans*-sensing. Aramayo and Metzenberg noticed that a cross between a dominant deletion mutant (*Asm-1*$^{\Delta}$ [ascus maturation]) and a wild-type strain (*Asm-1*$^{+}$) produces no viable ascospores. Interestingly, this sterile phenotype persists even with a cross between wild type and complemented *Asm-1*$^{\Delta}$ mutant with the gene located at an ectopic site (Aramayo et al. 1996). This dominant phenotype suggests that the silencing signal is *trans*-acting, as is the case in quelling. Indeed, genetic studies have identified a set of genes that are highly homologous to the components of the quelling pathway but that are specifically expressed in the meiotic stage. *S*uppressor of *a*scus *d*ominance 1 (*Sad-1*) is the first known meiotic silencing component identified by screening for mutants deficient in meiotic silencing; it encodes a putative RdRP homologous to *QDE-1* (Shiu et al. 2001). Recently, *Sad-3*, an RNA helicase homologous to fission yeast *Hrr1*, was shown to be involved in MSUD (Hammond et al. 2011). In the RNAi of fission yeasts, the RDRC complex consists of Rdp1 (an RdRP), *Cid12* (a polyA polymerase family member), and *Hrr1* (an RNA helicase). The identification of *Sad-1* and *Sad-3* indicated that the production of dsRNA via RdRP activity is an essential step in meiotic silencing as it is in quelling.

If the gene silencing effect of MSUD is mediated by an RNAi pathway, the dsRNA must be cleaved by Dicer and loaded onto Argonaute protein to cause silencing of genes complementary to the small RNA. Subsequent analyses identified a series of other MSUD-required genes that encode other RNAi components. *Neurospora* encodes two Dicer proteins, *Dcl-1* and *Dcl-2*. Although both DCLs have somewhat redundant functions in the quelling pathway, only *Dcl-1* (also called *Sms-3*) is required for MSUD, suggesting that *Dcl-1*/*Sms-3* but not *Dcl-2* is specifically expressed during meiosis (Alexander et al. 2008). Of the two identified Argonaute proteins encoded in the *Neurospora* genome, *QDE-2* and *S*uppressor of *m*eiotic *s*ilencing 2 (*Sms-2*), only the latter is required for MSUD, which is thought to function by binding to sRNAs originating from the unpaired DNA region.

Apart from these core RNAi components, other components are also required for MSUD. *Sad-2*, a fungus-specific protein with no apparent known domain structures, associates with *Sad-1* in vivo. *Sad-2* appears to function as a scaffold protein to recruit *Sad-1* to the perinuclear region (Shiu et al. 2006; Bardiya et al. 2008). Interestingly, *Dcl-1*/*Sms-3* and *Sms-2* are also localized in the perinuclear region, suggesting that this region is the major functional center for MSUD (Shiu et al. 2006; Alexander et al. 2008). In addition, QIP, an exonuclease required during the vegetative stage for the removal of passenger strands in quelling and maturation of milRNA (see below), is also involved in MSUD by locating at the perinuclear region (Lee et al. 2010a).

Since MSUD occurs in the diploid cells that account for a very small portion of perithecia and that exist in a short time window, it is difficult to use biochemical or northern analyses to detect the small RNA production from the MSUD pathway. By using high-throughput sequencing of small RNA from perithecia, Hammond et al. identified meiotic silencing-related siRNAs that likely associate with *Sms-2* (Hammond et al. 2013a). These sRNAs, called MSUD-associated small interfering RNAs or masiRNAs, are about 21–27 nt long (with a peak at 25 nt) and equally

match to both strands of unpaired DNA. Genetic and cell biological studies suggest that MSUD likely occurs through a two-step mechanism (Fig. 11.2). First, the *trans*-sensing mechanism scans for the presence or absence of paired homologous DNA. Second, the transcripts produced from unpaired genes are recognized by RNAi-related components specifically expressed during meiosis to trigger the production of siRNA and gene silencing. Most of the known MSUD components are localized mainly in the perinuclear region, suggesting that dsRNA and siRNA generation and silencing of the target mRNA mainly occurs there. It is unknown how the unpaired DNA is recognized in the nucleus. Recently, an MSUD mutant, *Sad-5*, was identified by a genetic screen (Hammond et al. 2013b). Interestingly, *Sad-5* is found in the nucleus. Future investigation of the role of *Sad-5* might offer insights into how the first step of MSUD, *trans*-sensing, occurs in the nucleus.

Diverse Biogenesis Pathways of miRNAs in *Neurospora*

miRNAs are found in plants, animals, and algae (Lee et al. 1993; Lagos-Quintana et al. 2001; Lee and Ambros 2001; Llave et al. 2002; Molnar et al. 2007; Zhao et al. 2007; Grimson et al. 2008). Unlike siRNAs, which are processing products of dsRNAs, the precursors of miRNAs are single-stranded RNAs that can form a hairpin structure (Ambros et al. 2003; Bartel 2004). miRNAs regulate many physiological and developmental processes by targeting mRNAs for degradation or translational repression (Bartel 2004). miRNAs were long considered to be absent from fungi.

Based on the definition of miRNA, it must meet three criteria: First, its precursor must be a single-stranded RNA that forms a hairpin structure. Second, the mature miRNA should be mostly derived from one stem of the hairpin. Third, miRNA should be able to silence the endogenous mRNA targets, either by resulting in mRNA degradation or translation repression. By analyzing the *QDE-2*-associated small RNAs in *N. crassa* by deep sequencing, 25 potential miRNA-producing loci were identified. The small RNAs produced from these loci satisfy all criteria listed above but were named miRNA-like small RNAs (milRNAs) to distinguish them from miRNAs found in plants and animals (Lee et al. 2010c).

In general, the miRNA biogenesis pathway consists of three steps: transcription of precursor RNAs, processing of the hairpin structure by Dicer, and formation of miRNA-containing RISC. Eukaryotic cells have three types of RNA polymerases: Pol I, which synthesizes 18S and 28S rRNA; Pol II, responsible for synthesis of mRNAs and most of miRNAs; and Pol III, which synthesizes 5S rRNAs, tRNAs, and some snRNAs. Unlike plant and animal miRNA genes that are mostly transcribed by Pol II, most *Neurospora* milRNAs are transcribed by RNA Pol III (Yang et al. 2013). Interestingly, even though inhibition of Pol II does not affect the synthesis of the most abundant milRNAs, Pol II only or both Pol II and Pol III were present at several examined milRNA-producing loci (Yang et al. 2013),

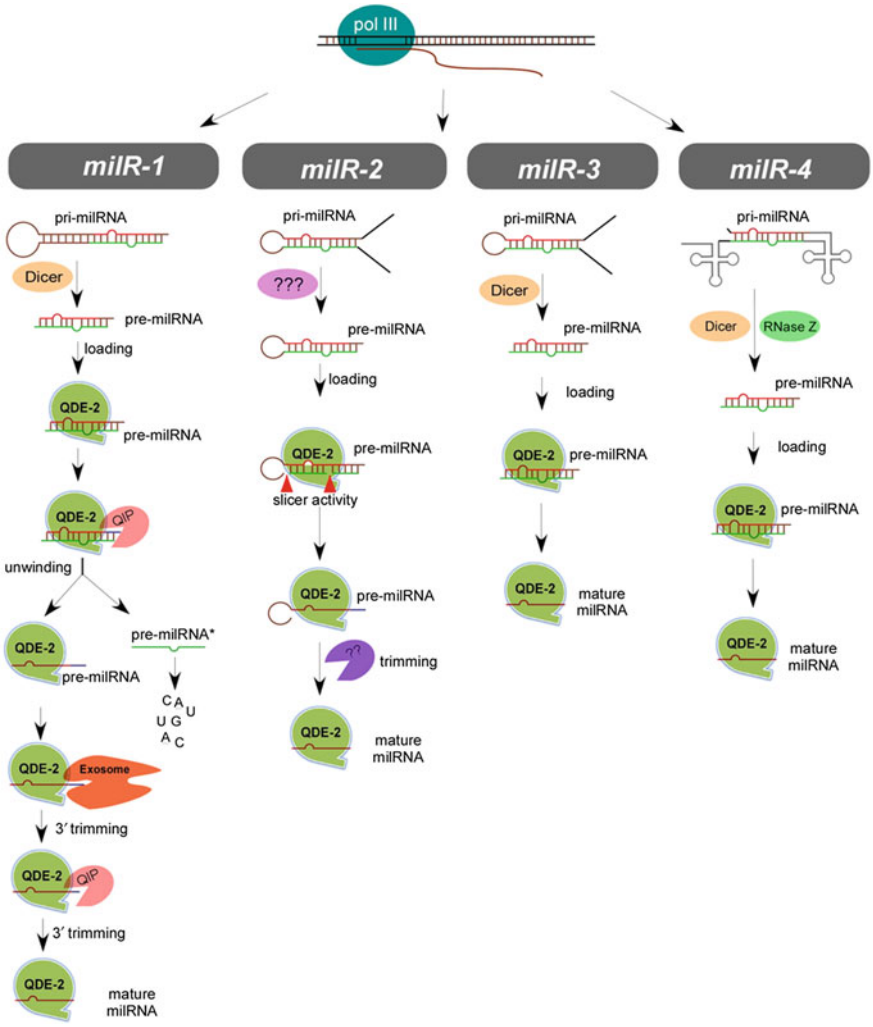


Fig. 11.3 The diverse pathways of microRNA-like sRNA (milRNA) biogenesis. From *left* to *right* are the processing pathways of four characterized milRNAs (*milR-1* to *-4*), all transcribed by RNA polymerase III. The putative structures of primary transcripts (*pri-milRNA*), precursors (*pre-milRNA*) and matured milRNA are depicted along the processing pathways

suggesting that Pol II and Pol III might coordinate to regulate the transcription of some milRNAs.

Surprisingly, milRNAs in *Neurospora* are produced by diverse biogenesis pathways (Lee et al. 2010c). Among the 25 identified milRNAs, four (*milR-1* to *-4*) have been well characterized (Fig. 11.3). The biogenesis of *milR-3* follows the canonical miRNA pathway: The hairpin *primary* (*pri*)*-milR-3* is processed by Dicer to produce mature milRNA, which is subsequently loaded onto Argonaute

protein *QDE-2*. Although the processing of *milR-4* is similar to that of *milR-3*, it is only partially dependent on Dicer. The *pri-milR-4* arises from a tRNA precursor. The production of *milR-4* requires RNase Z, the 3′ tRNA endonuclease (Yang et al. 2013). In contrast, the biogenesis of *milR-2* is completely independent of Dicer (Lee et al. 2010c). The precursor of *milR-2* (*pre-milR-2)* is predicted to form a hairpin structure, and RNA immunoprecipitation revealed that both mature *milR-2* and *pre-miR-2* are associated with *QDE-2*, but the maturation of the *pre-milR-2* does not rely on Dicer. Instead processing of *pre-milR-2* requires the slicer activity of *QDE-2*. These results suggest that *QDE-2* binds to the long *pre-milR-2* and cleaves the passenger strand. An unknown nuclease is then responsible for further cleavage and maturation of the milRNA. The *milR-2* biogenesis is the first known example of a Dicer-independent but Argonaute-dependent mechanism.

The *milR-1* milRNAs production pathway uses a novel 3′–5′ trimming mechanism for milRNA maturation and is currently the best understood (Lee et al. 2010c; Xue et al. 2012): First, the *pri-milR-1* is processed by Dicer into the *pre-milR-1* duplex. Second, the *pre-milR-1* duplex is loaded onto *QDE-2*. Third, QIP, a 3′–5′ exoribonuclease that is recruited by *QDE-2*, uses its helicase activity to separate the *pre-milR-1* duplex into ssRNAs in collaboration with *QDE-2*. The *pre-milR-1* strand remains on QDE-2 whereas the passenger strand is ejected from the complex and degraded. Fourth, the exosome trims the *QDE-2*-bound *pre-milR-1 s* from 3′ to 5′ end into sRNAs of intermediate sizes. Finally, the exosome-processed *pre-milR-1 s* are further processed into mature milRNAs in a process involving both QIP and exosome. In the biogenesis of *milR-1*, *QDE-2* plays three essential roles: It determines which strand of the *pre-milR-1* duplex is matured, recruits exosome and QIP, and determines the size of *milR-1* by protecting the mature *milR-1* from further processing.

The diversity of the milRNA biogenesis pathways in *Neurospora* offers important insights into eukaryotic small RNA biogenesis pathways. Soon after the discovery of *milR-2*, the mouse *miR-451* was shown to be produced by a Dicer-independent but Argonaute-dependent mechanism very similar to that of *milR-2* (Cheloufi et al. 2010; Cifuentes et al. 2010). Likewise, the biochemical demonstration of the *milR-1* biogenesis pathway and identification of exosome and QIP in miRNA maturation process offer important insights into how small RNAs can be matured by using a 3′–5′ trimming mechanism. It has been proposed that piRNA maturation requires a 3′–5′ trimming of the piRNA precursor bound to PIWI proteins (Kawaoka et al. 2011).

Do milRNAs in *Neurospora*, like those in plants and animals, regulate the gene expression? The answer is yes, at least for *milR-1*, as shown by three lines of evidence. First, the predicted target genes of *milR-1* are up-regulated in *Dcl* and *qde-2* mutants. Second, the protein levels produced from a reporter gene harboring a *milR-1* target site are dramatically elevated in the *qde-2* mutant. Third, predicted *milR-1*-targeted mRNAs are specifically associated with *QDE-2* (Lee et al. 2010c). Recently, milRNAs were discovered in *Cryptococcus neoformans* and *Penicillium marneffei* (Jiang et al. 2012; Lau et al. 2013). milRNAs do not appear to have crucial functions in growth or developmental processes under normal laboratory

conditions in *Neurospora* and *P. marneffei*. However, they may function under different growth conditions to fine-tune various biological processes. Indeed, *Botrytis cinerea*, an aggressive fungal pathogen, was recently found to silence expression of genes involved in immunity in *Arabidopsis* and tomato by hijacking the host Argonaute proteins with the fungal small RNAs (called Bc-sRNAs by the authors) that appear to be like milRNAs (Weiberg et al. 2013).

Dicer-Independent Small Interfering RNAs

Dicer-independent small interfering RNAs (disiRNAs) are another novel type of sRNA found in *Neurospora* (Lee et al. 2010c). Like qiRNAs and siRNAs, disiRNAs are also about 22-nt long and have a strong 5′ uridine preference, but the biogenesis of disiRNAs is independent of any known RNAi component including Dicer proteins. disiRNAs map to about 50 non-repetitive DNA loci and are nearly symmetrically distributed in both strands. The function of disiRNA is not well understood. However, disiRNAs arise from gene-rich loci, raising the possibility that they might be involved in regulation of gene expression. Indeed, many disiRNA-encoding loci are associated with DNA methylation, a mark of heterochromatin (Dang et al. 2013). The regulation of the disiRNA locus-linked DNA methylation (DLDM) is different from the well-studied DNA methylation from RIPped DNA regions. First, DLDM peaks in intergenic or promoter regions. Second, in the *disiRNA-47* loci that harbors the circadian clock gene *frequency*, DLDM fluctuates significantly depending on growth conditions (Belden et al. 2011). Third, the DLDM shows an on/off pattern (meaning that the locus is either densely methylated or un-methylated), suggesting that the methylation program is turned on only in some but not all nuclei (Dang et al. 2013). Fourth, DLDM requires transcription of the loci and importantly, convergent transcription can trigger DLDM at the promoter region (Belden et al. 2011; Dang et al. 2013). Consistent with this notion, most of the disiRNA loci are known to produce sense and antisense transcripts. However, because the biogenesis pathway of disiRNA is unknown, it is still unclear whether disiRNAs mediate DLDM.

Concluding Remarks

The studies of RNAi pathways in *Neurospora* in the past two decades have made fundamental contributions to the understanding of small RNA-mediated gene silencing mechanisms in eukaryotes. The discovery of diverse small RNA production pathways in *Neurospora* also offered important insights into how small RNAs are generated in eukaryotic organisms. Furthermore, *Neurospora* also serves a model system for the understanding of RNAi pathways and their functions in other fungal organisms (Dang et al. 2011; Chang et al. 2012; Billmyre et al. 2013).

Future studies in *Neurospora* on the quelling mechanism, miRNA biogenesis and functions, and other types of small RNAs will no doubt uncover important and novel insights into eukaryotic small RNA-mediated gene silencing mechanisms.

References

Alexander WG, Raju NB, Xiao H, Hammond TM, Perdue TD, Metzenberg RL, Pukkila PJ, Shiu PK (2008) DCL-1 colocalizes with other components of the MSUD machinery and is required for silencing. Fungal Genet Biol 45:719–727

Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-Jones S, Marshall M et al (2003) A uniform system for microRNA annotation. RNA 9:277–279

Aramayo R, Metzenberg RL (1996) Meiotic transvection in fungi. Cell 86:103–113

Aramayo R, Peleg Y, Addison R, Metzenberg R (1996) Asm-1+, a *Neurospora crassa* gene related to transcriptional regulators of fungal development. Genetics 144:991–1003

Bardiya N, Alexander WG, Perdue TD, Barry EG, Metzenberg RL, Pukkila PJ, Shiu PK (2008) Characterization of interactions between and among components of the meiotic silencing by unpaired DNA machinery in *Neurospora crassa* using bimolecular fluorescence complementation. Genetics 178:593–596

Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116:281–297

Belden WJ, Lewis ZA, Selker EU, Loros JJ, Dunlap JC (2011) CHD1 remodels chromatin and influences transient DNA methylation at the clock gene frequency. PLoS Genet 7:e1002166

Bernstein E, Caudy AA, Hammond SM, Hannon GJ (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature 409:363–366

Billmyre RB, Calo S, Feretzaki M, Wang X, Heitman J (2013) RNAi function, diversity, and loss in the fungal kingdom. Chromosome Res. doi:10.1007/s10577-10013-19388-10572

Bzymek M, Lovett ST (2001) Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. Proc Natl Acad Sci USA 98:8319–8325

Calzada A, Hodgson B, Kanemaki M, Bueno A, Labib K (2005) Molecular anatomy and regulation of a stable replisome at a paused eukaryotic DNA replication fork. Genes Dev 19: 1905–1919

Catalanotto C, Azzalin G, Macino G, Cogoni C (2000) Gene silencing in worms and fungi. Nature 404:245

Catalanotto C, Azzalin G, Macino G, Cogoni C (2002) Involvement of small RNAs and role of the *Qde* genes in the gene silencing pathway in *Neurospora*. Genes Dev 16:790–795

Catalanotto C, Pallotta M, ReFalo P, Sachs MS, Vayssie L, Macino G, Cogoni C (2004) Redundancy of the two dicer genes in transgene-induced posttranscriptional gene silencing in *Neurospora crassa*. Mol Cell Biol 24:2536–2545

Chang SS, Zhang Z, Liu Y (2012) RNA interference pathways in fungi: mechanisms and functions. Annu Rev Microbiol 66:305–323

Cheloufi S, Dos Santos CO, Chong MM, Hannon GJ (2010) A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. Nature 465:584–589

Chen H, Samadder PP, Tanaka Y, Ohira T, Okuizumi H, Yamaoka N, Miyao A, Hirochika H, Tsuchimoto S, Ohtsubo H et al (2008) *OsRecQ1*, a *Qde-3* homologue in rice, is required for RNA silencing induced by particle bombardment for inverted repeat DNA, but not for double-stranded RNA. Plant J 56:274–286

Cifuentes D, Xue H, Taylor DW, Patnode H, Mishima Y, Cheloufi S, Ma E, Mane S, Hannon GJ, Lawson ND et al (2010) A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. Science 328:1694–1698

Cogoni C, Irelan JT, Schumacher M, Schmidhauser TJ, Selker EU, Macino G (1996) Transgene silencing of the al-1 gene in vegetative cells of *Neurospora* is mediated by a cytoplasmic effector and does not depend on DNA–DNA interactions or DNA methylation. EMBO J 15:3153–3163

Cogoni C, Macino G (1997) Isolation of quelling-defective (*Qde*) mutants impaired in posttranscriptional transgene-induced gene silencing in *Neurospora crassa*. Proc Natl Acad Sci USA 94:10233–10238

Cogoni C, Macino G (1999a) Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase. Nature 399:166–169

Cogoni C, Macino G (1999b) Posttranscriptional gene silencing in *Neurospora* by a RecQ DNA helicase. Science 286:2342–2344

Cogoni C, Romano N, Macino G (1994) Suppression of gene expression by homologous transgenes. Antonie Van Leeuwenhoek 65:205–209

Colot HV, Park G, Turner GE, Ringelberg C, Crew CM, Litvinkova L, Weiss RL, Borkovich KA, Dunlap JC (2006) A high-throughput gene knockout procedure for *Neurospora* reveals functions for multiple transcription factors. Proc Natl Acad Sci USA 103:10352–10357

Covey SN, AlKaff NS, Langara A, Turner DS (1997) Plants combat infection by gene silencing. Nature 385:781–782

Crowley TE, Nellen W, Gomer RH, Firtel RA (1985) Phenocopy of discoidin I-minus mutants by antisense transformation in Dictyostelium. Cell 43:633–641

Dang Y, Li L, Guo W, Xue Z, Liu Y (2013) Convergent transcription induces dynamic DNA methylation at disiRNA loci. PLoS Genet 9:e1003761

Dang Y, Yang Q, Xue Z, Liu Y (2011) RNA interference in fungi: pathways, functions, and applications. Eukaryot Cell 10:1148–1155

Djupedal I, Portoso M, Spahr H, Bonilla C, Gustafsson CM, Allshire RC, Ekwall K (2005) RNA Pol II subunit *Rpb7* promotes centromeric transcription and RNAi-directed chromatin silencing. Genes Dev 19:2301–2306

Ecker JR, Davis RW (1986) Inhibition of gene expression in plant cells by expression of antisense RNA. Proc Natl Acad Sci USA 83:5372–5376

Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature 411:494–498

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 391:806–811

Francia S, Michelini F, Saxena A, Tang D, de Hoon M, Anelli V, Mione M, Carninci P, d'Adda di Fagagna F (2012) Site-specific DICER and DROSHA RNA products control the DNA-damage response. Nature 488: 231–235

Freitag M, Lee DW, Kothe GO, Pratt RJ, Aramayo R, Selker EU (2004) DNA methylation is independent of RNA interference in *Neurospora*. Science 304:1939

Fulci V, Macino G (2007) Quelling: post-transcriptional gene silencing guided by small RNAs in *Neurospora crassa*. Curr Opin Microbiol 10:199–203

Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma LJ, Smirnov S, Purcell S et al (2003) The genome sequence of the filamentous fungus *Neurospora crassa*. Nature 422:859–868

Galagan JE, Selker EU (2004) RIP: the evolutionary cost of genome defense. Trends Genet 20:417–423

Grimson A, Srivastava M, Fahey B, Woodcroft BJ, Chiang HR, King N, Degnan BM, Rokhsar DS, Bartel DP (2008) Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature 455:1193–1197

Hamilton AJ, Baulcombe DC (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. Science 286:950–952

Hammond SM, Boettcher S, Caudy AA, Kobayashi R, Hannon GJ (2001) Argonaute2, a link between genetic and biochemical analyses of RNAi. Science 293:1146–1150

Hammond TM, Spollen WG, Decker LM, Blake SM, Springer GK, Shiu PK (2013a) Identification of small RNAs associated with meiotic silencing by unpaired DNA. Genetics 194:279–284

Hammond TM, Xiao H, Boone EC, Decker LM, Lee SA, Perdue TD, Pukkila PJ, Shiu PK (2013b) Novel proteins required for meiotic silencing by unpaired DNA and siRNA generation in *Neurospora crassa*. Genetics 194:91–100

Hammond TM, Xiao H, Boone EC, Perdue TD, Pukkila PJ, Shiu PK (2011) *Sad-3*, a putative helicase required for meiotic silencing by unpaired DNA interacts with other components of the silencing machinery. G3 1:369–376

Harris SD (2001) Septum formation in *Aspergillus nidulans*. Curr Opin Microbiol 4:736–739

Hsieh J, Fire A (2000) Recognition and silencing of repeated DNA. Annu Rev Genet 34:187–204

Huang J, Brito IL, Villen J, Gygi SP, Amon A, Moazed D (2006) Inhibition of homologous recombination by a cohesin-associated clamp complex recruited to the rDNA recombination enhancer. Genes Dev 20:2887–2901

Jiang N, Yang Y, Janbon G, Pan J, Zhu X (2012) Identification and functional demonstration of miRNAs in the fungus *Cryptococcus neoformans*. PLoS ONE 7:e52734

Kato A, Akamatsu Y, Sakuraba Y, Inoue H (2004) The *Neurospora crassa Mus-19* gene is identical to the *Qde-3* gene, which encodes a RecQ homologue and is involved in recombination repair and postreplication repair. Curr Genet 45:37–44

Kawaoka S, Izumi N, Katsuma S, Tomari Y (2011) 3' end formation of PIWI-interacting RNAs in vitro. Mol Cell 43:1015–1022

Kim SK, Wold BJ (1985) Stable reduction of thymidine kinase activity in cells expressing high levels of anti-sense RNA. Cell 42:129–138

Kinsey JA (1990) Tad, a line-like transposable element of *Neurospora*, can transpose between nuclei in heterokaryons. Genetics 126:317–323

Krings G, Bastia D (2004) swi1- and swi3-dependent and independent replication fork arrest at the ribosomal DNA of *Schizosaccharomyces pombe*. Proc Natl Acad Sci USA 101:14085–14090

Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T (2001) Identification of novel genes coding for small expressed RNAs. Science 294:853–858

Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, Kingston RE (2006) Characterization of the piRNA complex from rat testes. Science 313:363–367

Lau SKP, Chow WN, Wong AYP, Yeung JMY, Bao J, Zhang N, Lok S, Woo PCY, Yuen KY (2013) Identification of MicroRNA-Like RNAs in Mycelial and yeast phases of the thermal dimorphic fungus *Penicillium marneffei*. Plos Neglected Trop Dis 7:e2398

Lee DW, Millimaki R, Aramayo R (2010a) QIP, a component of the vegetative RNA silencing pathway, is essential for meiosis and suppresses meiotic silencing in *Neurospora crassa*. Genetics 186:127–133

Lee HC, Aalto AP, Yang Q, Chang SS, Huang G, Fisher D, Cha J, Poranen MM, Bamford DH, Liu Y (2010b) The DNA/RNA-dependent RNA polymerase *Qde-1* generates aberrant RNA and dsRNA for RNAi in a process requiring replication protein A and a DNA helicase. PLoS Biol 8:e1000496

Lee HC, Chang SS, Choudhary S, Aalto AP, Maiti M, Bamford DH, Liu Y (2009) qiRNA is a new type of small interfering RNA induced by DNA damage. Nature 459:274–277

Lee HC, Li L, Gu W, Xue Z, Crosthwaite SK, Pertsemlidis A, Lewis ZA, Freitag M, Selker EU, Mello CC et al (2010c) Diverse pathways generate microRNA-like RNAs and Dicer-independent small interfering RNAs in fungi. Mol Cell 38:803–814

Lee RC, Ambros V (2001) An extensive class of small RNAs in *Caenorhabditis elegans*. Science 294:862–864

Lee RC, Feinbaum RL, Ambros V (1993) The *C. elegans* heterochronic gene *Lin-4* encodes small RNAs with antisense complementarity to *Lin-14*. Cell 75:843–854

Li L, S-S Chang, Liu Y (2010a) RNA interference pathways in filamentous fungi. Cell Mol Life Sci 67:3849–3863

Li L, Chang SS, Liu Y (2010b) RNA interference pathways in filamentous fungi. Cell Mol Life Sci 67:3849–3863

Liu Y, Ye X, Jiang F, Liang C, Chen D, Peng J, Kinch LN, Grishin NV, Liu Q (2009) C3PO, an endoribonuclease that promotes RNAi by facilitating RISC activation. Science 325:750–753

Llave C, Kasschau KD, Rector MA, Carrington JC (2002) Endogenous and silencing-associated small RNAs in plants. Plant Cell 14:1605–1619

Luteijn MJ, Ketting RF (2013) PIWI-interacting RNAs: from generation to transgenerational epigenetics. Nat Rev Genet 14:523–534

Maiti M, Lee HC, Liu Y (2007) QIP, a putative exonuclease, interacts with the *Neurospora* Argonaute protein and facilitates conversion of duplex siRNA into single strands. Genes Dev 21:590–600

Michalik KM, Bottcher R, Forstemann K (2012) A small RNA response at DNA ends in *Drosophila*. Nucleic Acids Res 40:9596–9603

Mohanty BK, Bairwa NK, Bastia D (2006) The Tof1p-Csm3p protein complex counteracts the Rrm3p helicase to control replication termination of *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA 103:897–902

Molnar A, Schwach F, Studholme DJ, Thuenemann EC, Baulcombe DC (2007) miRNAs control gene expression in the single-cell alga *Chlamydomonas reinhardtii*. Nature 447:1126–1129

Nakanishi K, Weinberg DE, Bartel DP, Patel DJ (2012) Structure of yeast Argonaute with guide RNA. Nature 486:368–374

Napoli C, Lemieux C, Jorgensen R (1990) Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous Genes in trans. Plant Cell 2:279–289

Nolan T, Braccini L, Azzalin G, De Toni A, Macino G, Cogoni C (2005) The post-transcriptional gene silencing machinery functions independently of DNA methylation to repress a *Line1*-like retrotransposon in *Neurospora crassa*. Nucleic Acids Res 33:1564–1573

Nolan T, Cecere G, Mancone C, Alonzi T, Tripodi M, Catalanotto C, Cogoni C (2008) The RNA-dependent RNA polymerase essential for post-transcriptional gene silencing in *Neurospora crassa* interacts with replication protein A. Nucleic Acids Res 36:532–538

Pickford A, Braccini L, Macino G, Cogoni C (2003) The *Qde-3* homologue *RecQ-2* co-operates with *Qde-3* in DNA repair in *Neurospora crassa*. Curr Genet 42:220–227

Raju NB (1980) Meiosis and ascospore genesis in *Neurospora*. Eur J Cell Biol 23:208–223

Raju NB (1992) Genetic-control of the sexual cycle in *Neurospora*. Mycol Res 96:241–262

Ratcliff F, Harrison BD, Baulcombe DC (1997) A similarity between viral defense and gene silencing in plants. Science 276:1558–1560

Romano N, Macino G (1992) Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences. Mol Microbiol 6:3343–3353

Salgado PS, Koivunen MR, Makeyev EV, Bamford DH, Stuart DI, Grimes JM (2006) The structure of an RNAi polymerase links RNA silencing and transcription. PLoS Biol 4:e434

Schramke V, Sheedy DM, Denli AM, Bonila C, Ekwall K, Hannon GJ, Allshire RC (2005) RNA-interference-directed chromatin modification coupled to RNA polymerase II transcription. Nature 435:1275–1279

Selker EU (2002) Repeat-induced gene silencing in fungi. Adv Genet 46:439–450

Sharma V, Misteli T (2013) Non-coding RNAs in DNA damage and repair. FEBS Lett 587:1832–1839

Shiu PK, Raju NB, Zickler D, Metzenberg RL (2001) Meiotic silencing by unpaired DNA. Cell 107:905–916

Shiu PK, Zickler D, Raju NB, Ruprich-Robert G, Metzenberg RL (2006) *Sad-2* is required for meiotic silencing by unpaired DNA and perinuclear localization of *Sad-1* RNA-directed RNA polymerase. Proc Nat Acad Sci U S A 103:2243–2248

Vader G, Blitzblau HG, Tame MA, Falk JE, Curtin L, Hochwagen A (2011) Protection of repetitive DNA borders from self-induced meiotic instability. Nature 477:115–119

Van Blokland R, Van der Geest N, Mol JNM, Kooter JM (1994) Transgene-mediated suppression of chalcone synthase expression in Petunia hybrida results from an increase in RNA turnover. Plant J 6:861–877

Verdel A, Jia S, Gerber S, Sugiyama T, Gygi S, Grewal SI, Moazed D (2004) RNAi-mediated targeting of heterochromatin by the RITS complex. Science 303:672–676

Wang YL, Juranek S, Li HT, Sheng G, Tuschl T, Patel DJ (2008a) Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. Nature 456:U921–U972

Wang YL, Sheng G, Juranek S, Tuschl T, Patel DJ (2008b) Structure of the guide-strand-containing argonaute silencing complex. Nature 456:U209–U234

Wei W, Ba Z, Gao M, Wu Y, Ma Y, Amiard S, White CI, Rendtlew Danielsen JM, Yang YG, Qi Y (2012) A role for small RNAs in DNA double-strand break repair. Cell 149:101–112

Weiberg A, Wang M, Lin FM, Zhao HW, Zhang ZH, Kaloshian I, Huang HD, Jin HL (2013) Fungal small RNAs suppress plant immunity by Hijacking host RNA interference pathways. Science 342:118–123

Wierzbicki AT, Haag JR, Pikaard CS (2008) Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. Cell 135:635–648

Xue Z, Yuan H, Guo J, Liu Y (2012) Reconstitution of an argonaute-dependent small RNA biogenesis pathway Reveals a handover mechanism involving the RNA exosome and the exonuclease QIP. Mol cell 46:299–310

Yang Q, Li L, Xue Z, Ye Q, Zhang L, Li S, Liu Y (2013) Transcription of the major *Neurospora crassa* microRNA-like small RNAs relies on RNA polymerase III. PLoS Genet 9:e1003227

Zamore PD, Tuschl T, Sharp PA, Bartel DP (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell 101:25–33

Zhang Z, Chang SS, Zhang Z, Xue Z, Zhang H, Li S, Liu Y (2013) Homologous recombination as a mechanism to recognize repetitive DNA sequences in an RNAi pathway. Genes Dev 27:145–150

Zhao T, Li G, Mi S, Li S, Hannon GJ, Wang XJ, Qi Y (2007) A complex system of small RNAs in the unicellular green alga *Chlamydomonas reinhardtii*. Genes Dev 21:1190–1203

Chapter 12
The RNAi Machinery in Mucorales: The Emerging Role of Endogenous Small RNAs

Victoriano Garre, Francisco E. Nicolás, Santiago Torres-Martínez and Rosa M. Ruiz-Vázquez

Abstract The increasing knowledge on the functional relevance of endogenous small RNAs as riboregulators has stimulated the identification and characterization of these molecules in many eukaryotes. The small RNA molecules act to suppress gene expression through RNA interference (RNAi) pathways, which may differ among organisms. These pathways are normally involved in genome defense and heterochromatin formation, while their role in endogenous gene regulation in fungal models has been scarcely studied. *Mucor circinelloides* is a basal fungus of the subphylum Mucoromycotina, evolutionary distant to Ascomycota and Basidiomycota fungi. *M. circinelloides* has been used as model system for studying the silencing mechanism and the role of this mechanism in the regulation of endogenous functions. This review summarizes our current knowledge on the genes involved in the *M. circinelloides* RNAi pathway, whose central core consists of Dicer, Argonaute, and RNA-dependent RNA polymerase proteins. We will also discuss the possible functionality of the RNAi machinery in other Mucoromycotina species, given the presence of the core components of this machinery in their genome sequences. Finally, we will present an overview of the different classes of endogenous small RNAs accumulated by *M. circinelloides* and how these RNAs may regulate gene expression and control different cellular functions.

V. Garre · F. E. Nicolás · S. Torres-Martínez · R. M. Ruiz-Vázquez (✉)
Department of Genetics and Microbiology, University of Murcia, Murcia, Spain
e-mail: rmruiz@um.es

V. Garre
e-mail: vgarre@um.es

F. E. Nicolás
e-mail: fnicolas@um.es

S. Torres-Martínez
e-mail: storres@um.es

A. Sesma and T. von der Haar (eds.), *Fungal RNA Biology*,
DOI: 10.1007/978-3-319-05687-6_12,

Contents

Discovery of Gene Silencing in *Mucor*

Gene silencing was the initial hypothesis proposed to explain an apparent paradox observed early in the 2000s in *M. circinelloides,* a basal fungus that is a causal agent for the rare but lethal fungal infection mucormycosis, an emerging infectious disease recognized as a prevalent fungal infection in patients with impaired immunity (Ribes et al. 2000). *M. circinelloides f. lusitanicus* belongs to the family Mucoraceae, order Mucorales, subphylum Mucoromycotina (Hibbett et al. 2007) and it has been extensively used as model organism for studying light responses, such as the blue light activation of carotene biosynthesis (reviewed by Ruiz-Vázquez and Torres-Martínez 2003). During a search for genes involved in the carotenogenesis pathway, our group isolated the *crgA* gene, which was identified as an ORF which causes carotene overaccumulation in the dark when introduced as a transgene in the wild-type strain of *M. circinelloides* (Navarro et al. 2000). Surprisingly, lack of *crgA* function also provoked the same phenotype of carotenoid overaccumulation both in the dark and the light (Navarro et al. 2001). These results resembled the phenomenon designated as "quelling" (see Chap. 11), discovered by Romano and Macino (1992) in *Neurospora crassa* after transforming a wild-type strain with different portions of the carotenogenic albino-3 (*al-3*) and albino-1 (*al-1*) genes, which encode the geranylgeranyl pyrophosphate synthase and the phytoene dehydrogenase, respectively. They observed that, instead of enhancing gene expression by increasing gene dosage, transformation with these transgenes caused silencing of the duplicated genes. Thus, we proposed that the apparently contradictory *crgA* results could be explained if a silencing-like mechanism of gene expression were operative in the *crgA* multicopy transformants, resulting in the suppression of *crgA* protein synthesis

(Navarro et al. 2001). Later on, it was demonstrated that "quelling" is a posttranscriptional gene silencing mechanism (PTGS), similar to RNA interference (RNAi), an evolutionarily conserved mechanism for silencing gene expression.

The existence of a gene silencing mechanism was later demonstrated in *M. circinelloides* by using a simple visual reporter system to analyze transgene-induced gene silencing (Nicolás et al. 2003). Wild-type strains transformed with complete or truncated copies of the *carB* gene, coding for the phytoene dehydrogenase enzyme, presented an albino phenotype instead of the bright yellow color seen when the *carB* gene is expressed at the wild-type levels. The albino phenotype is a consequence of a strong reduction in the steady-state levels of mature (spliced) *carB* mRNA, whereas unspliced mRNA levels were the same in wild-type and albino transformants. This established that transgene-induced gene silencing in *M. circinelloides* occurs at a posttranscriptional level, thus providing a solid support to solve the *crgA* paradox.

Mucor as a Model Organism to Study Gene Silencing

M. circinelloides has become a good model organism for the study of different molecular processes in the fungal kingdom, including light responses, pathogenesis, and gene silencing, mainly due to the availability of a large number of molecular tools and to its evolutionary distance from other fungal model organisms, such as *N. crassa.* An added value of the genetic transformation system of *M. circinelloides* is that, unlike most ascomycetes and basidiomycetes, transforming DNA does not integrate into the genome but is maintained in an episomal state. Thus, transgene expression is not affected by position effects or host regulatory sequences at insertion sites, both of which are thought to be involved in the production of abnormally processed RNAs. These aberrant RNAs could be used as substrates for RNA-dependent RNA polymerases (RdRPs) to synthesize double-stranded RNA (dsRNA), the molecules that trigger the silencing mechanism (Fire et al. 1998). The nonintegrative nature of *M. circinelloides* transgenes allowed us to demonstrate that a high level of expression of transgenes is essential for silencing, since the silencing frequency increased from 3 to 90 % when the expression of the transgene was increased in different ways (Nicolás et al. 2009; de Haro et al. 2009; Calo et al. 2012). This could be explained if a high transgene expression results in the accumulation of enough aberrant RNA to activate RdRP-mediated copying.

The canonical RNA-silencing mechanism, or RNA interference (RNAi), is triggered by dsRNA molecules and results in the production of small interfering RNA (siRNAs) molecules of defined sizes, which act as effectors for the suppression of gene expression. Production of siRNAs is mediated by the Dicer protein, a member of the RNaseIII family of dsRNA-specific endonucleases, and suppression of gene expression is mediated by an RNA-induced silencing complex

(RISC), which can recognize and cleave a target RNA complementary to the guide strand of the siRNA (reviewed in Carthew and Sontheimer 2009; Ghildiyal and Zamore 2009). The Argonaute (AGO) protein is the core component of the RISC complex. *M. circinelloides* follows this canonical mechanism with some specific peculiarities, since it is associated with the production of two size classes of antisense siRNAs, 21- and 25-nt long, which are differentially accumulated during vegetative growth (Nicolás et al. 2003). The biological function of these two different classes of siRNAs and the differential roles that they could play during growth and development are not fully understood, although several evidences suggest a central role of the 21-nt siRNA in degradation of the target mRNA and silencing efficiency and stability (Nicolás et al. 2009).

Amplification of Silencing

In plants and nematodes, the RNAi mechanism is supplemented through the action of an RdRP activity that expands the initial siRNA production (primary siRNAs) with the generation of secondary siRNAs. Secondary siRNAs, contrary to primary siRNAs, do not derive from the initial triggering molecule but they come from the targeted mRNAs, which are used as templates by the RdRP proteins. While primary siRNAs only correspond to sequences of the dsRNA molecules that initiate the silencing mechanism, secondary siRNAs may also correspond to sequences of the endogenous gene upstream and downstream of the initial inducer sequence. The amplification process differs in plant and nematodes. In plants, the 5′ and 3′ fragments of mRNAs that have been targeted by primary siRNAs and cleaved by Argonaute are recognized as aberrant RNA by an RdRP protein, which synthesizes a complementary strand to produce new dsRNA molecules. This dsRNA is then processed by Dicer into secondary siRNAs with the typical 5′ monophosphate ends. In the nematode *Caenorhabditis elegans*, target mRNAs bound to primary siRNA-Argonaute complexes are recognized by an RdRP protein that directly synthesize short unprimed antisense siRNA in a Dicer-independent manner. As a consequence, these secondary siRNAs contain a triphosphate group at the 5′ end and they are predominantly antisense to corresponding mRNAs (reviewed in Ghildiyal and Zamore 2009).

The amplification process is also present in *M. circinelloides* (Nicolás et al. 2003), which most probably uses a Dicer-dependent mechanism to produce secondary siRNAs. Although *rdrp* genes have been identified in several fungi, the role of those genes in amplification of the silencing signal through the production of secondary siRNAs has been scarcely analyzed. In *N. crassa* and *Aspergillus nidulans*, the *rdrp* genes are not required for gene silencing when using dsRNA as inducer and no secondary siRNAs have been detected, suggesting the absence of a siRNA amplification step mediated by RdRPs (Chang et al. 2012). Unlike those fungi, in *M. circinelloides* both size classes of siRNAs are present as secondary siRNA (Nicolás et al. 2003). The 25-nt class includes sense and antisense

molecules, which suggests that these secondary siRNAs derive from dicing of dsRNAs generated from RdRP-dependent conversion of the targeted mRNA, as it occurs in plants. On the other hand, 21-nt siRNAs are mainly antisense to the target mRNA, which could suggest that they are directly synthesized by an RdRP using the target mRNA as a template, as it happens in *C. elegans*. However, 21-nt antisense siRNAs have 5′ monophosphate ends (see Sect. "The RNA-Dependent RNA Polymerases"), which suggests that they also derive from dicing of dsRNA. In plants, secondary siRNAs can be detected from sequences upstream and downstream of the initial triggering molecule, indicating bidirectional spreading from the initiator region into adjacent regions of the target gene (Ghildiyal and Zamore 2009). The spreading of silencing in *M. circinelloides* operates mostly in the 5′ to 3′ direction, since the two size classes of antisense siRNAs were preferentially produced from sequences of the target gene downstream of the input trigger. This suggests that the uncapped 3′ cleavage fragments are preferentially recognized as aberrant RNA by the *M. circinelloides* RdRP proteins. Unlike other organisms where the same RdRP is responsible for the induction of silencing by sense transgenes and the amplification of secondary siRNAs, in *M. circinelloides* distinct RdRPs are involved in these two processes (Calo et al. 2012) (see Sect. "The RNA-Dependent RNA Polymerases").

RNAi Machinery in *M. circinelloides*

RNAi is an evolutionary conserved defense mechanism to fight against exogenous nucleic acids such as viruses, transposons, and transgenes. Besides that, the RNAi pathways regulate gene expression in the vast majority of eukaryotic organisms, controlling growth and differentiation and helping to maintain genome integrity (see Sect. "Endogenous Small RNAs in *M. circinelloides*"). However, differences exist in the specific RNAi pathways that operate in distinct organisms and in the nature and number of the proteins involved (Cerutti and Casa-Mollano 2006). In the fungal kingdom, these pathways entail three specialized elements that are the central core of the RNAi machinery: an RdRP protein, the Dicer enzyme, and the Argonaute protein. These three elements have been identified and thoroughly analyzed in the fungus *M. circinelloides*.

The Dicer Enzyme

The Dicer enzyme belongs to the RNaseIII family of ribonucleases, which specifically cleave dsRNA giving rise to small duplexes with a monophosphate group at the 5′ end and two nucleotide overhangs on their 3′ end. The Dicer enzyme is found only in eukaryotes and displays a complex structure: two tandemly arranged RNaseIII domains and a single dsRNA-binding domain (dsRBD) at the carboxyl

terminus, an amino-terminal ATP-dependent RNA helicase domain (DEXD/H-box domain), a small domain of unknown function proposed to fold as a dsRNA-binding domain (the DUF283 domain), and a Piwi-Argonaute-Zwille (PAZ) domain, which binds specifically to the 3′ end of ssRNA (Jaskiewicz and Filipowicz 2008). The distance between the PAZ domain and the catalytic RNaseIII domains, which is determined by a nonconserved connector helix, controls the length of the Dicer products. Thus, the presence of two different size classes of antisense siRNAs in *M. circinelloides* could suggest the involvement of two different Dicer proteins in the RNAi pathway of this fungus.

The first RNAi machinery element that was cloned and analyzed in *M. circinelloides* was a *dicer-like* (*dcl*) gene (Nicolás et al. 2007). The *dcl-1* gene product contained all the structural domains normally found in other proteins of the Dicer family, including the RNA helicase DEXD/H-box domain, the PAZ domain, two catalytic RNaseIII domains, and a C-terminal dsRNA-binding domain. Null mutants for the *dcl-1* gene were not impaired in the transgene-induced silencing mechanism, since they presented identical silencing frequency as the wild-type strain. These mutants were also able to produce the two size classes of siRNAs, 21- and 25-nt long, when silencing was induced by both sense and dsRNA-producing transgenes (Nicolás et al. 2007; de Haro et al. 2009). This demonstrated that the *dcl-1* gene is not essential for the production of any of the two classes of siRNAs associated with transgene-induced silencing in *M. circinelloides* and implied that, at least, one additional *dicer* gene would have to exist in this fungus to produce the siRNA molecules. Surprisingly, phenotypic analyses of the *dcl-1*$^{-}$ mutants revealed that these strains showed a reduction in their vegetative growth rate and presented an altered hyphal morphology, which suggested a role of *dcl-1* in the regulation of endogenous cellular processes (Nicolás et al. 2007). These were the first evidences implying that the RNA interference mechanism in fungi, besides its defense function, could have a role in development and physiology. Later, phenotypic alterations in a few filamentous fungi have been reported for mutants in the RNAi machinery components (Nakayashiki and Nguyen 2008; Alexander et al. 2008; de Haro et al. 2009; Cervantes et al. 2013; Carreras-Villaseñor et al. 2013), although in most cases no phenotypic changes have been observed.

M. circinelloides has a second *dicer-like* gene, *dcl-2*, which plays a pivotal role in the production of siRNAs triggered by both sense and dsRNA-producing transgenes, as gene silencing is severely impaired in the *dcl-2*$^{-}$ mutant (Fig. 12.1) (de Haro et al. 2009). These findings contrast with results in *N. crassa*, where the two *dicer* genes are functionally redundant for the silencing function (Catalanotto et al. 2004). The *M. circinelloides dcl-2*$^{-}$ mutant does not accumulate any of the two size classes of antisense siRNAs, suggesting that both are generated by the same Dcl-2 activity (de Haro et al. 2009). This also occurs with the ascomycete *Magnaporthe oryzae*, where only one of the two *dicer* genes is involved in the production of different size classes of siRNAs (Kadotani et al. 2004). Thus, the situation in a number of fungi seems to be divergent from that described in plants, where distinct Dicer enzymes are responsible for the generation of different size siRNAs that act in different silencing pathways (Xie et al. 2004).

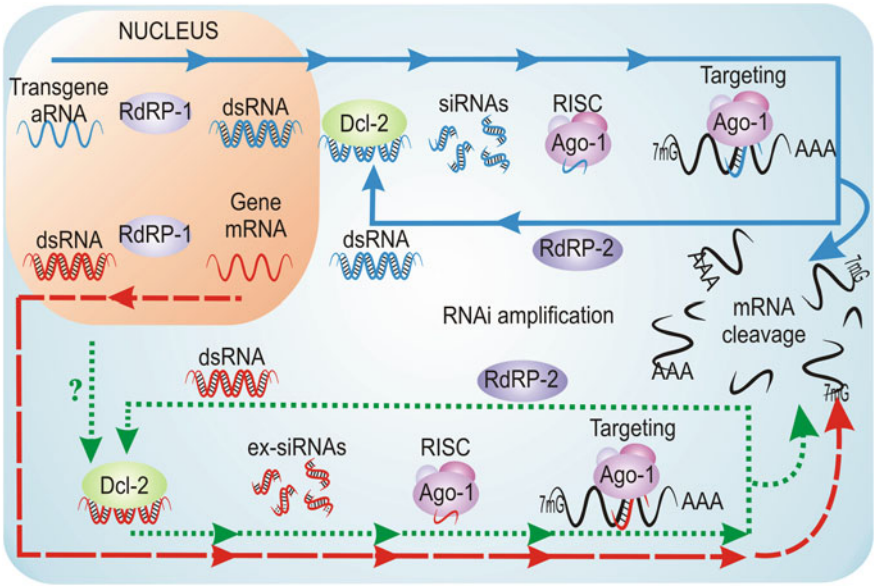


Fig. 12.1 Two different RNAi pathways in *M. circinelloides*. The transgene-induced RNAi pathway (*solid line*) is activated by aberrant RNAs (aRNA) derived from the transgenes, which are converted to dsRNA precursors by RdRP-1. These dsRNAs are processed by Dcl-2 into siRNAs that are loaded onto Ago-1-based RISC for cleaving the target RNAs. The resultant RNA fragments become the substrate for RdRP-2, which generates new dsRNA that is incorporated to the RNAi pathway at the dicing stage. The processing of this new dsRNA produces secondary siRNAs that amplify the silencing signal. The endogenous regulatory RNAi pathway produces different classes of ex-siRNAs derived from endogenous mRNAs. The main class of these ex-siRNAs (class 2) is produced by a canonical RNAi pathway (*dashed line*) that requires RdRP-1 to convert mRNA derived from selected loci into dsRNAs, which are processed by Dcl-2 into ex-siRNAs that specifically bind to Ago-1 to suppress the expression of the corresponding target genes. In contrast, class 1 ex-siRNAs require Dcl-2, Ago-1, and RdRP-2, but not RdRP-1, for their biogenesis (*dotted line*). Biogenesis of ex-siRNAs of classes 3 and 4 is not shown

The *M. circinelloides dcl-2*$^{-}$ mutant exhibits a large reduction in the silencing frequency, but not a complete loss (de Haro et al. 2009). This residual activity, which is not sufficient to maintain the silenced phenotype, is due to the Dicer-1 protein, since gene silencing is completely abolished in the *dcl-1*$^{-}$/*dcl-2*$^{-}$ double mutant. Thus, although *dcl-2* plays the main role in transgene-induced gene silencing, a partial redundancy with *dcl-1* exists. Both the *dcl-2*$^{-}$ and *dcl-1*$^{-}$/*dcl-2*$^{-}$ mutants are affected in vegetative development, since they show a significant reduction in the production of asexual spores (de Haro et al. 2009). They also show an accelerated autolytic response to nutritional stress (Cervantes et al. 2013), a process that is not a simple cell necrosis phenomenon but is an active and well-regulated process where many enzymatic activities are involved (Emri et al. 2008). These results point to a crucial role for *dcl-2* in an endogenous gene regulation mechanism in *M. circinelloides*.

The Argonaute Protein

Argonaute proteins are the core components of RISC, a multiprotein complex that functions as the final effector in the silencing pathway by targeting the mRNA through complementarity with the single-stranded siRNA. The AGO protein family is characterized by the presence of PAZ, MID (Middle), and PIWI domains (Wei et al. 2012). The PIWI domain is the catalytic center of AGO, which possesses a cleaving activity similar to RNase H. Within this domain is the catalytic triad DDH, in which the two aspartate residues are highly conserved whereas histidine can be substituted by other residues.

AGO proteins are conserved between species, although they have undergone remarkable structural evolution and functional diversification (Wei et al. 2012). The number of *argonaute* genes identified in filamentous fungal genomes range from zero to as many as eight (Nakayashiki et al. 2006). In *M. circinelloides*, three different *ago* genes have been identified and thoroughly characterized (Cervantes et al. 2013). The three *ago* genes code for highly similar proteins, which contain all the structural domains characteristics of the AGO family, including the catalytic triad DDD. However, functional analysis of the corresponding null mutants revealed that only the *ago-1* gene plays a key role in the RNAi mechanism of *M. circinelloides* during vegetative growth (Fig. 12.1). Neither *ago-2* nor *ago-3* is involved in vegetative gene silencing and the corresponding mutants do not have recognizable phenotypes, although a possible role for these genes cannot be discarded at different growth conditions or specialized structures. The *ago-1* gene is required for transgene-induced silencing whatever the nature of the silencing trigger is, as the corresponding mutant shows a negative silencing phenotype for both sense and inverted repeat transgenes. Contrary to what happens in other fungi, such as *N. crassa* (Catalanotto et al. 2002), Ago-1 is also required for production and/or stability of siRNAs, since no siRNAs have been detected in the *ago-1*$^{-}$ mutant (Cervantes et al. 2013). It has been proposed that Ago-1 could have a role in the biogenesis of secondary siRNAs, as occurs in metazoan, although lack of siRNA accumulation in *ago-1*$^{-}$ mutants could also indicate that these small RNAs are stabilized by binding to Ago-1, so that in its absence the siRNAs would be rapidly degraded (Cervantes et al. 2013).

The essential role of the *ago-1* gene in the RNAi pathway explains the phenotype observed in the *ago-1*$^{-}$ mutant. As *dcl-2*$^{-}$, the *ago-1*$^{-}$ mutant presents a strong reduction in the production of asexual spores and an accelerated autolytic response to nutritional stress (Cervantes et al. 2013). Sporulation and autolysis are developmental responses in which many genes are involved, showing the importance of the RNAi machinery in the endogenous regulation of complex developmental processes (discussed in Sect. "Endogenous Small RNAs in *M. circinelloides*").

The RNA-Dependent RNA Polymerases

Dicer and Argonaute proteins are two essential components of the RNAi machinery that can be found across the domain eukarya. However, the RdRP component is only found in plants, nematodes, and some fungi, such as *M. circinelloides*, where it confers the capability of amplifying the silencing signal through the production of secondary siRNAs. Besides enhancing the RNAi response, the RdRP proteins also function in the initiation of silencing by sense transgenes, through the production of the triggering dsRNA molecules from single-stranded transcripts derived from the transgenes. In most organisms, the same enzyme participates in both processes, being absolutely required to initiate silencing by sense transgenes but affecting only the stability of the silenced phenotype when silencing is induced by dsRNA molecules. Recently, a third function has been proposed for RdRPs in the RNAi mechanism, namely the direct production of aberrant RNA from a DNA template (Lee et al. 2010). The synthesis of the aberrant RNA has been traditionally assigned to the RNA polymerase II, which would transcribe the ssDNA template to produce the aberrant RNA used by the RdRPs to generate dsRNA. However, the *N. crassa* QDE-1 protein, which is an RdRP enzyme, also has DNA-dependent RNA polymerase (DdRP) activity. Thus, this RdRP could be required for both the synthesis of the aberrant RNA and the subsequent dsRNA production.

There are three *rdrp* genes in *M. circinelloides*. Whereas *rdrp-3* does not have any role in the transgene-induced silencing mechanism during the vegetative growth (our unpublished results), RdRP-1 and RdRP-2 proteins play differential roles in the initiation and amplification steps of RNAi (Calo et al. 2012) (Fig. 12.1). RdRP-1 is required for initiation of silencing by sense transgenes but not for the accumulation of secondary siRNAs when silencing is triggered by dsRNA-producing transgenes. The *rdrp-1*$^{-}$ mutant is unable to produce antisense RNA transcripts derived from sense transgenes, which confirms the role of the *rdrp-1* gene in the conversion of ssRNA transcripts into dsRNA molecules. RdRP-1 plays the same role in *M. circinelloides* than QDE-1 in *N. crassa*, which is essential for initiation of silencing by sense transgenes but it is not required for efficient gene silencing when it is triggered by dsRNA-expressing constructs (Chang et al. 2012). However, the DdRP activity of the *M. circinelloides* RdRP-1 protein has not yet been demonstrated. On the other hand, *M. circinelloides* RdRP-2 is only involved in the amplification step of the silencing mechanism, being required for the efficient accumulation of the two size classes of secondary siRNAs (Calo et al. 2012). The *rdrp-2*$^{-}$ mutants show a very low silencing frequency and very unstable silenced phenotype when silencing is induced by both sense and inverted repeat transgenes. However, these mutants are perfectly able to produce antisense RNA molecules from sense transgenes, indicating that *rdrp-2* is not involved in initiation of silencing by sense transgenes. The differentiated roles of RdRP-1 and RdRP-2 in the silencing mechanism indicate a functional diversification of these proteins, which have evolved to participate in different steps of the same RNA silencing pathway.

The fact that the RdRP-2 enzyme is required for the production of both the secondary sense and antisense 25-nt siRNAs and the strictly antisense 21-nt siRNA class, could suggest that this enzyme is responsible for the generation of long dsRNA molecules from single-stranded transcripts, as well as for the direct synthesis of 21-nt antisense siRNAs using the target mRNA as a template. However, sequencing of secondary siRNAs corresponding to the reporter gene *carB* demonstrated that both the antisense 21-nt and the sense and antisense 25-nt siRNAs contain 5′ monophosphate ends, suggesting that the two classes derive from dicing dsRNAs (our unpublished results). Thus, it is possible that hypothetical interactions of RdRP-2 with Dcl-2 could polarize the processing activity of Dicer on newly synthesized dsRNA, allowing selective stabilization of the antisense strand after cleavage of dsRNA, as proposed in other organisms (Lee and Collins 2007). If this hypothesis is correct, the differential accumulation of the two size classes of siRNAs during the vegetative growth could be due to the specific regulation of the interactions between RdRP-2 and Dcl-2 proteins during the vegetative cycle (Calo et al. 2012).

Evolutionary Conservation of RNAi Machinery in Mucorales

Conservation of the RNAi machinery based on database searches for RdRP-, Argonaute-, and Dicer-like proteins in genome sequences is a valuable alternative to knowing the existence of the RNAi pathway in an organism, provided that genetic or biochemical strategies are not available. Comparative sequence analyses suggest that the RNAi machinery is well conserved in the eukaryote kingdom, indicating that it appeared before the divergence of major eukaryotic phylogenetic lineages (Nakayashiki et al. 2006). Interestingly, some particular species of fungi and protozoa have lost this machinery during evolution (Nicolás et al. 2013). However, although loss or attenuation of the RNAi pathway may give a selective advantage to some of these species, the deleterious effects of active transposons and the lack of defenses against new evolving viruses might condemn RNAi-deficient species to extinction over a long evolutionary term (Nicolás et al. 2013).

M. circinelloides f. lusitanicus is the only basal fungus where the existence of RNAi pathways has been proven experimentally. A wider presence of the RNAi pathway in Mucorales is supported by the identification of hypothetical genes coding for the whole RNAi protein set in genomes of different mucoralean families. Especially, this set has been found in *Rhizopus delemar* (a cryptic species of the *R. oryzae* complex), which belongs to the family Rhizopodaceae, and *Phycomyces blakesleeanus*, which belongs to the family Phycomyceteaceae (our unpublished results). Moreover, the presence of this machinery has been also found in a strain of *M. circinelloides f. circinelloides* isolated from skin samples from a healthy human volunteer. The search for RNAi components in other

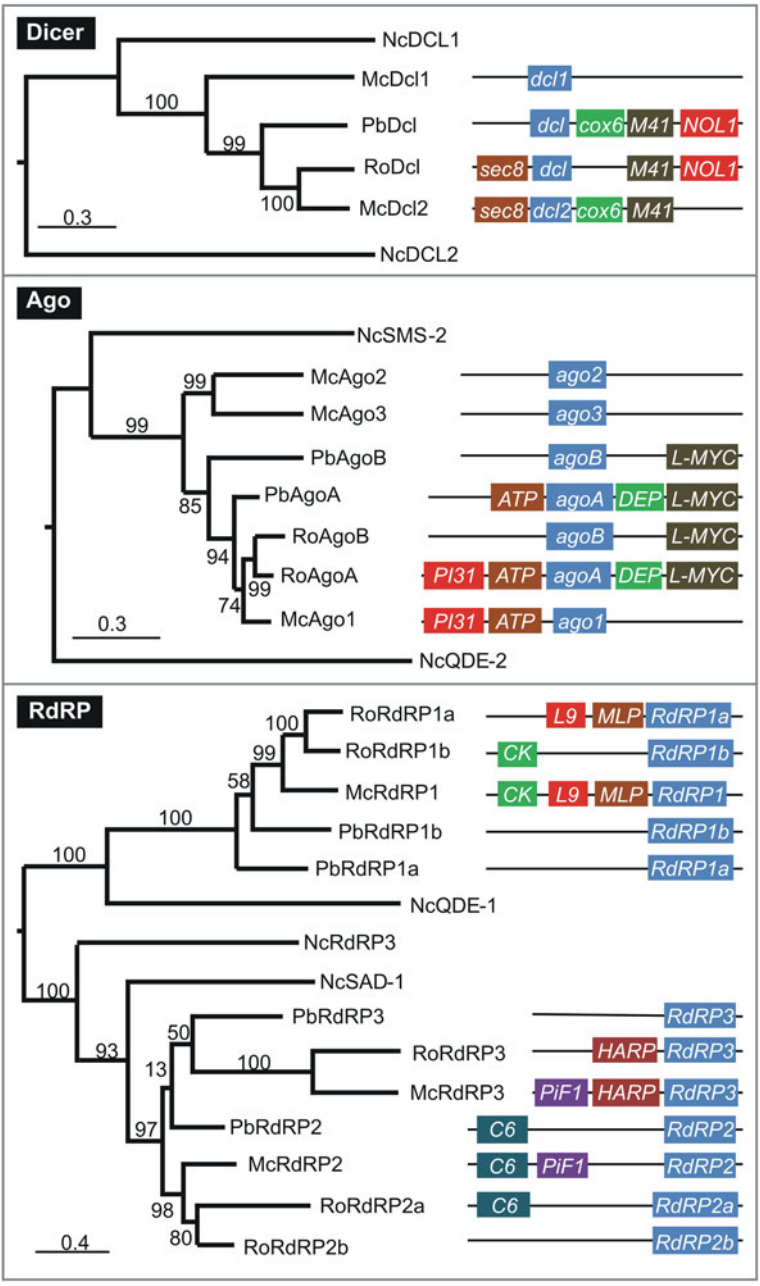
Mucoromycotina species without sequenced genomes has been unsuccessful, probably due to the low number of available Mucoromycotina sequences deposited in GenBank (Hoffmann et al. 2013).

To gain information about the functionality of genes coding for RNAi proteins identified in database searching, additional features were analyzed, including protein domain structure, phylogenetic relations based on protein sequence similarity, synteny, and gene expression data, when available. Expression of RNAi genes in *R. delemar* and *P. blakesleeanus* have not been reported, but unpublished expression data are available in FungiDB for *R. delemar* (Stajich et al. 2012) and in its genome website (http://genome.jgi.doe.gov/Phybl2/Phybl2.home.html) for *P. blakesleeanus*. Those websites include RNAseq data representing different times of hyphal growth for *R. delemar*, and EST data from dark and light grown mycelia for *P. blakesleeanus*, respectively. Conservation analysis of the RNAi components in Mucorales revealed a copy number variability for each component in the analyzed species (Fig. 12.2), probably as a result of copy loss or duplications after species divergence. A description of the phylogenetic relationships and feasible functionality for each RNAi protein in Mucorales is shown below.

Dicer Proteins

Mucoralean Dicer proteins are phylogenetically close to *N. crassa* DCL1, suggesting that they probably derived from a common ancestor, whereas proteins derived from the *N. crassa* DCL2 ancestor were lost during Mucoralean evolution (Fig. 12.2). The two *M. circinelloides* species have two *dcl* genes, while *R. delemar* and *P. blakesleeanus* have only one. Sequence similarity and synteny analyses suggested that the single Dicer enzymes of *R. delemar* and *P. blakesleeanus* are phylogenetically close to *M. circinelloides* Dcl-2, which is the Dicer enzyme involved in RNA silencing (de Haro et al. 2009; Nicolas et al. 2010). Moreover, they maintain the typical domain structure for Dicer proteins, except the PAZ domain in the *R. delemar* protein. However, the absence of this domain does not preclude the functionality of this protein, because the functional *N. crassa* DCLs also lack this domain (Catalanotto et al. 2004). In addition, expression of the *R. delemar* gene has been detected in vegetative mycelium. Therefore, it is probable that *R. delemar* and *P. blakesleeanus* Dicer proteins are functional, having a similar role to *M. circinelloides* Dcl-2.

Dcl-1 proteins are specific for *M. circinelloides*, although the duplication from which the *dcl-1* gene originated probably occurred before the divergence of *M. circinelloides* from the rest of Mucorales (Fig. 12.2), since a truncated *dcl-1-like* gene was found in *R. delemar* genome. Therefore, Dcl-1 proteins may carry out functions that are particular for *M. circinelloides*, but absent in closely related Mucorales.


Dicer
NcDCL1
McDcl1
dcl1
100
PbDcl
dcl
cox6
M41
NOL1
99
RoDcl
sec8
dcl
M41
NOL1
100
McDcl2
sec8
dcl2
cox6
M41
0.3
NcDCL2
Ago
NcSMS-2
99
McAgo2
ago2
McAgo3
ago3
99
PbAgoB
agoB
L-MYC
85
PbAgoA
ATP
agoA
DEP
L-MYC
RoAgoB
agoB
L-MYC
94
RoAgoA
PI31
ATP
agoA
DEP
L-MYC
99
74
McAgo1
PI31
ATP
ago1
0.3
NcQDE-2
RdRP
100
RoRdRP1a
L9
MLP
RdRP1a
99
RoRdRP1b
CK
RdRP1b
58
McRdRP1
CK
L9
MLP
RdRP1
100
PbRdRP1b
RdRP1b
100
PbRdRP1a
RdRP1a
NcQDE-1
NcRdRP3
NcSAD-1
100
PbRdRP3
RdRP3
50
93
RoRdRP3
HARP
RdRP3
13
100
McRdRP3
PiF1
HARP
RdRP3
97
PbRdRP2
C6
RdRP2
McRdRP2
C6
PiF1
RdRP2
98
RoRdRP2a
C6
RdRP2a
80
RoRdRP2b
RdRP2b
0.4

◀ **Fig. 12.2** Phylogenetic relationships of Mucoralean silencing proteins (Dicer, Argonaute and RdRP). Phylogenetic trees for silencing proteins of *R. delemar* (Ro), *M. circinelloides f. lusitanicus* (Mc), *P. blakesleeanus* (Pb), and *N. crassa* (Nc), used as an outgroup species, were constructed as previously described (Calo et al. 2012). Branch lengths are proportional to the number of substitutions per site (bars). The numbers at the nodes are bootstrap values (%) for 100 replications. Synteny among silencing genes is shown on the *right side* of the figure. Only syntenic genes relevant for phylogenetic analysis are shown. Genes and distances are not at scale. Synteny was determined using FungiDB (FungiDB.org) and manual inspection of the *Phycomyces* genome. Accession numbers of silencing proteins for *M. circinelloides* and *P. blakesleeanus* at JGI (Grigoriev et al. 2011), and *N. crassa* and *R. delemar* at Broad Institute of Harvard and MIT (http://www.broadinstitute.org/): McDcl1, 104148; Mcdcl2, 104153; Pbdcl, 85799; NcDCL1, NCU08270; NcDCL2, NCU06766; RoDcl, RO3G_15434; PbAgoA, 123569; PbAgoB, 85795; McAgo1, 104161; McAgo2, 195366; McAgo3, 104163; RoAgoB, RO3G_10137; RoAgoA, RO3G_13047; NcQDE-2, NCU04730; NcSMS-2, NCU09434; PbRdRP1a, 59322; PbRdRP1b, 65575; PbRdRP2, 85804; PbRdRP3, 85806; NcQDE-1, NCU07534; NcSAD-1, NCU02178; NcRdRP3, NCU08435; RoRdRP1a, RO3G_08583; RoRdRP1b, RO3G_11872; RoRdRP2a, RO3G_11051; RoRdRP3, RO3G_10258; RoRdRP2b, RO3G_14914; McRdRP1, 111871; McRdRP2, 195368; McRdRP3, 159162. Gene abbreviations: *sec8*, exocyst complex subunit Sec8; *cox6*, Cytochrome c oxidase subunit 6; *M41*, AAA+-type ATPase containing the peptidase M41 domain; *NOL*1, nucleolar protein NOL1; *L-MYC*, L-myc-2 protein; *ATP*, calcium transporting ATPase; *DEP*, protein containing DEP domain; *PI31*, proteasome formation inhibitor PI31; *L9*, 60S ribosomal protein L9; *MLP*, regulatory protein MLP; *CK*, checkpoint kinase; *HARP*, chromatin remodeling protein HARP/SMARCAL1; *PiF1*, DNA helicase PIF1/RRM3; *C6*, C6 finger domain protein

Argonaute Proteins

Protein similarity analyses suggested that Mucoralean Argonaute proteins are phylogenetically closer to SMS-2 than QDE-2, two *N. crassa* Argonaute proteins involved in different RNA pathways (Fig. 12.2). This was unexpected, since SMS-2 is essential for meiotic silencing by unpaired DNA (MSUD), an RNA pathway that has been found only in ascomycetes, whereas QDE-2 is required for vegetative silencing (Chang et al. 2012). *M. circinelloides f. lusitanicus* has three different *ago* genes, but only *ago-1* is required for RNA silencing during vegetative growth (Cervantes et al. 2013). The *M. circinelloides ago* genes derive from two duplication events, the common ancestor of *ago-2* and *ago-3* genes being generated in the same duplication as the *ago-1* ancestor (Fig. 12.2). Orthologs of *ago-2* and *ago-3* genes were only found in *M. circinelloides f. circinelloides*. The two *ago* genes found in *R. delemar* and *P. blakesleeanus* derived from the last common *ago-1* ancestor and were probably originated in recent segmental duplications after species divergence, according to sequence similarity and synteny (Fig. 12.2). The *R. delemar* and *P. blakesleeanus* Ago proteins show the typical domain architecture of the Ago family, suggesting that they are functional. This is further supported by expression data, since expression of the *agoA* genes of *P. blakesleeanus* and *R. delemar* has been detected, although expression of the *agoB* genes was not clearly sustained by the available data.

RdRP Proteins

M. circinelloides f. lusitanicus has three different *rdrp* genes, which play different functions (see Sect. "The RNA-Dependent RNA Polymerases"). Similarly, *M. circinelloides f. circinelloides* has also three genes that are phylogenetically related to each *M. circinelloides f. lusitanicus rdrp* gene. A comparative genomic approach revealed that orthologs of each *M. circinelloides rdrp* gene are present in the analyzed Mucorales, although segmental duplications have increased the numbers of *rdrp* genes in *R. delemar* and *P. blakesleeanus* relative to *M. circinelloides* (Fig. 12.2). Thus, *R. delemar* has five hypothetical *rdrp* genes, two of which are phylogenetically related to *rdrp-1*, two are related to *rdrp-2* and one is an *rdrp-3* ortholog. Remarkably, the duplicated genes of *R. delemar* and *P. blakesleeanus* correspond to *rdrp* genes with known functions in the *M. circinelloides* silencing pathway (see Sect. "The RNA-Dependent RNA Polymerases"). In addition, all Mucoralean RdRPs contain a typical RdRP domain, suggesting that they can be functional. This is further supported by the expression data for all *rdrp* genes of *R. delemar*. Expression of *P. blakesleeanus rdrp* genes could not be clearly determined with the available data.

Comparison among Mucoralean and *N. crassa* RdRPs revealed that they clustered in two different groups, each including *N. crassa* proteins, which suggests that they derived from two ancestors present prior to the divergence of Mucoromycotina and Ascomycotina. Excitingly, *M. circinelloides* RdRP-1 and *N. crassa* QDE-1, which play the same role in the silencing mechanism (see Sect. "The RNA-Dependent RNA Polymerases"), appeared in the same cluster (Fig. 12.2), suggesting that RdRP-1 orthologs of other Mucorales carry out the same function. Both Mucoralean RdRP-2 and RdRP-3 are close to *N. crassa* SAD-1, pointing out that those Mucoralean RdRPs are the result of a duplication event occurred in Mucorales prior speciation (Fig. 12.2). Despite their phylogenetic relationship, *M. circinelloides* RdRP-2 and RdRP-3 have acquired different functions, since RdRP-2 is involved in amplification of silencing during vegetative growth whereas no role is known for RdRP-3 in vegetative silencing.

The genomic comparative approach described in this section indicates that *R. delemar* and *P. blakesleeanus*, in addition to *M. circinelloides f. circinelloides*, have all the main components of the RNAi machinery, suggesting that this machinery is functional. Although orthologs of *dcl-1*, *ago-2* and *ago-3* genes are not present in the *R. delemar* and *P. blakesleeanus* genomes, these Mucorales contain, at least, one copy of the pivotal genes required for transgene-induced silencing, as well as for endogenous silencing (see Sect. "Endogenous Small RNAs in *M. circinelloides*"), suggesting that both processes occur in both fungi. The fact that *M. circinelloides*, *R. delemar* and *P. blakesleeanus* belong to distinct Mucoralean families, points out that RNAi pathways must be present in other Mucorales. A complementary conclusion is that *M. circinelloides dcl-1*, *ago-2* and *ago-3* could perform some *M. circinelloides* specific functions, not yet characterized.

Endogenous Small RNAs in *M. circinelloides*

Although the silencing machinery seems to be present in all Mucorales investigated, *M. circinelloides f. lusitanicus* is the only fungus of this class in which the RNAi pathway has been dissected and the main genes involved have been identified and functionally characterized. *M. circinelloides* mutants affected in the silencing mechanism showed phenotypic changes relative to the wild-type strain (see Sect. "RNAi Machinery in *M. circinelloides*), which suggested a role for the silencing machinery in the regulation of endogenous functions. In fact, analysis of endogenous small RNAs (esRNAs) in *M. circinelloides* revealed a plethora of regulatory esRNA molecules that control gene expression (Nicolás et al. 2010; Cervantes et al. 2013).

Sequencing analysis of small RNAs (sRNAs) accumulated in the wild-type strain and mutants affected in silencing genes allowed the identification of about a thousand of esRNA producing loci, which were *dicer*-dependent, since they showed a significant decrease in esRNA accumulation in *dcl-1*$^{-}$ or/and *dcl-2*$^{-}$ mutants relative to the wild type (Nicolás et al. 2010). These esRNA loci corresponded to repetitive sequences and transposons, intergenic regions and exons. Surprisingly, the analysis of the distribution of these esRNAs among the different types of loci revealed that they were not formed randomly across the genome, but they were enriched in exonic sequences compared with intergenic and repetitive regions. This is in contrast with other fungi, such as *Schizosaccharomyces pombe* (Grewal and Jia 2007), where esRNAs correspond to centromeric repeats and guide heterochromatin formation, *Saccharomyces castellii* and *Candida albicans* (Drinnenberg et al. 2009), where most esRNAs are produced from repeats and transposons, and *N. crassa*, where they mainly derive from highly repetitive rDNA loci (Lee et al. 2009) and intergenic regions (Lee et al. 2010). Sequencing data indicated that Dcl-2 is the primary protein involved in the production of all classes of esRNAs in *M. circinelloides*, since more than 84 % of the esRNA producing loci exclusively depend on Dcl-2 for their biogenesis (Fig. 12.1).

A small number of miRNA-like sRNAs (milRNAs) have been described in several ascomycetes and basidiomycetes, such as *N. crassa* (Lee et al. 2010), *Cryptococcus neoformans* (Jiang et al. 2012), *Sclerotinia sclerotiorum* (Zhou et al. 2012) and *Penicillium marneffei* (Lau et al. 2013), although the physiological role of these milRNAs is still unknown. In most cases, milRNA precursors were coded in intergenic regions and they adopted the stem-loop structure characteristic of miRNA loci. Although a significant number of *M. circinelloides* esRNAs derive from intergenic and intronic loci, none of the *dicer*-dependent loci corresponding to those regions fulfilled the criteria of *bona fide* miRNA loci. Extension of the analysis to all esRNA loci initially identified, including those that were not downregulated in *dcl*$^{-}$ mutants, confirmed that none of them had the features of miRNA genes. Thus, it seems that miRNAs are not present in *M. circinelloides*, at least under the growth conditions used in the experiment (Nicolás et al. 2010).

Table 12.1 Characteristics of the four classes of ex-siRNAs (modified from Nicolás et al. 2010; Cervantes et al. 2013)

Class	No. of exons	Strand bias[a]	Downregulated in	Binding to Ago-1	5′ U[b] (%)
1	9	−0.78	*dcl-2*$^{-}$, *rdrp-2*$^{-}$, *ago-1*$^{-}$	+	92.18
2	222	−0.34	*dcl-2*$^{-}$, *rdrp-1*$^{-}$, *ago-1*$^{-}$	+	92.12
3	88	0.90	*dcl-1*$^{-}$/*dcl-2*$^{-}$, *rdrp-1*$^{-}$, *rdrp-2*$^{-}$, *ago-1*$^{-}$	–	8.39
4	5	0.83	*dcl-1*$^{-}$, *rdrp-1*$^{-}$, *rdrp-2*$^{-}$, *ago-1*$^{-}$	–	28.28

[a] Strand bias indicates orientation to mRNAs, where 1 corresponds to all sRNAs in the same orientation as the mRNA, 0 to equal mixture of sRNAs on both strands and −1 to all sRNAs antisense to mRNAs

[b] The percentage of redundant reads that contain an uracil in the 5′ end of the molecule

ex-siRNAs

The exonic-siRNAs (ex-siRNAs) are the main class of *M. circinelloides* esRNAs (Nicolás et al. 2010). Sequencing analysis identified 324 ex-siRNA producing loci that corresponded to 276 genes, since some genes contain more that one ex-siRNAs producing exon. The ex-siRNAs regulate the expression of the protein-coding genes from which they were produced by guiding degradation of the corresponding mRNAs. Thus, the mRNAs of the target genes accumulated at a high level in those silencing mutants that are unable to produce the corresponding ex-siRNAs. There are four different classes of ex-siRNAs (classes 1–4) that have been classified based on their structural characteristics and the differential silencing machinery involved in their biogenesis (Table 12.1). Classes 1 and 2 correspond to those ex-siRNAs that depend on Dcl-2 and Ago-1 for their production (Nicolás et al. 2010; Cervantes et al. 2013). They present a defined size of 23–24 nt and show a strong preference for uracil at the 5′ end of the molecule. Class 2 ex-siRNAs also require RdRP-1 but not RdRP-2 for their biogenesis. The requirement of RdRP-1 and Dcl-2 for class 2 ex-siRNA biogenesis suggests that mRNAs from those loci are converted into dsRNA by RdRP-1 and then processed by Dcl-2 (Fig. 12.1). The involvement of these two proteins in the biogenesis of the major class of ex-siRNAs can be extended to esRNAs derived from transposons and intergenic regions, since most of the Dcl-2-dependent loci derived from those regions also require RdRP-1 (Nicolás et al. 2010). Only a small number of Dcl-2-dependent ex-siRNAs do not require RdRP-1 for their biogenesis but most of them depend on RdRP-2. This is the signature of class 1 ex-siRNAs (Table 12.1), which otherwise have structural characteristics similar to that of the class 2. The isolation of Ago-1-bound esRNAs from wild-type *M. circinelloides* confirmed that ex-siRNAs of classes 1 and 2 bind specifically to Ago-1 (Table 12.1), which indicates that they are functional siRNAs produced by a canonical RNAi pathway to suppress the expression of the corresponding target genes (Cervantes et al. 2013).

Class 3 of ex-siRNAs corresponds to those that present reduced accumulation only in the double *dcl-1*$^{-}$/*dcl-2*$^{-}$ mutant but not in *dcl-1*$^{-}$ or *dcl-2*$^{-}$ single mutants, indicating a redundant function of the two *dicer* genes in the production of these ex-siRNAs (Nicolás et al. 2010). This class, which covers a significant group of ex-siRNAs (Table 12.1), also requires the RdRP-1, RdRP-2, and Ago-1 proteins, as the levels of these ex-siRNAs are downregulated in the corresponding mutants (Nicolás et al. 2010; Cervantes et al. 2013). Besides the requirement of a different combination of silencing proteins for their biogenesis, class 3 ex-siRNAs show differential structural features relative to classes 1 and 2. Thus, ex-siRNAs of class 3 display a random spread of size distribution and they do not present a preference for Uracil in the 5′ end of the molecules. Besides that, they show a very strong strand bias, almost all of them being exclusively sense to the mRNAs, as expected from a degradation process. This has led to the suggestion that they are not generated by a canonical silencing pathway (Nicolás et al. 2010). This noncanonical biogenesis pathway would also require the Ago-1 protein, as downregulation of class 3 ex-siRNAs in the *ago-1*$^{-}$ mutant cannot be explained by stabilization of these ex-siRNAs by Ago binding, since they are not detected among the Ago-1-bound ex-siRNAs (Table 12.1) (Cervantes et al. 2013). The preference for 5′ uracil of *M. circinelloides* Ago-1 may explain lack of binding of class 3 ex-siRNAs, which show a preference for adenine at the 5′ end. Although these ex-siRNAs could bind to a different *M. circinelloides* Ago protein, the unusual structural characteristics of class 3 ex-siRNAs point to a noncanonical biogenesis mechanism, a frequent situation in filamentous fungi (Lee et al. 2010). It has been proposed that the sequential or combined activity of the *M. circinelloides* RdRP proteins could generate short dsRNA with target sites that are distributed along the target mRNA. These discrete dsRNA regions would be processed by either Dcl-1 or Dcl-2 and, after the initial cleavage, the single-stranded portions of mRNAs would be degraded by nonspecific RNases, probably because they lose their cap and/or poly-A tail (Nicolás et al. 2010). Ago-1 would participate, together with RdRP and Dicer proteins, in the multiprotein complex that targets the mRNAs that has to be degraded to an unidentified ribonuclease (Cervantes et al. 2013).

The relationship between the mechanisms that control mRNA quality and the RNAi machinery is increasingly evident, not only by the participation of the exosome in the biogenesis of esRNAs (Xue et al. 2012), but also by the cooperation between both mechanisms for binding and processing aberrant mRNAs (Yamanaka et al. 2013). In higher eukaryotes, oversaturation of the mRNA quality control mechanisms may result in triggering of gene silencing to degrade aberrant mRNAs. It is possible that, in basal fungi, the basic components of the silencing machinery might be used to channel aberrant mRNAs for degradation by unspecific nucleases or for processing into canonical esRNAs. How the *M. circinelloides* silencing components discriminate what RNAs are directed to the canonical silencing pathway or to the degradation pathway is not known. However, it is possible that the level of gene transcription plays an important role in the outcome of aberrant mRNAs. In silico analysis of genes regulated by class 3 ex-siRNAs

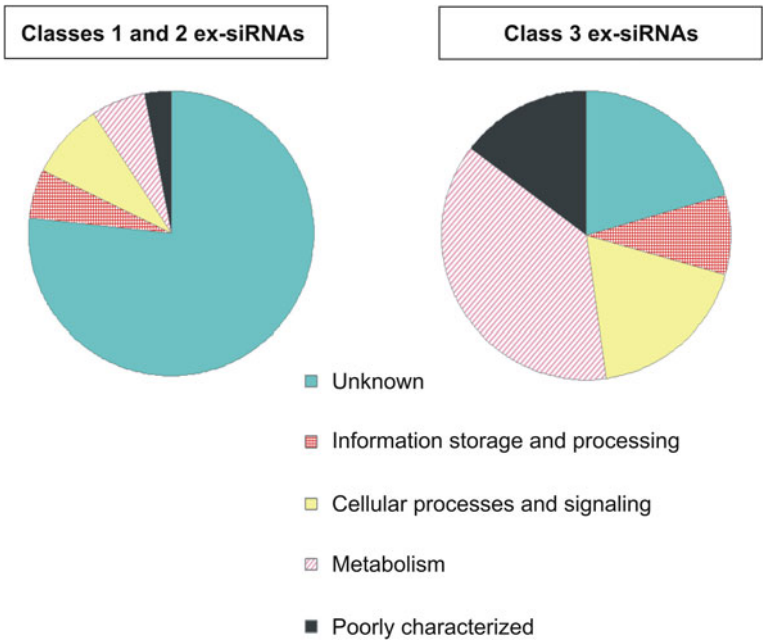


Fig. 12.3 Pie chart distribution of *M. circinelloides* ex-siRNA loci among different functional categories. ex-siRNAs of classes 1 and 2 (*left*) and ex-siRNAs of class 3 (*right*) were assigned to different functional categories based on KOG (EuKaryotic Orthologous Groups) database

reveals that they are highly expressed genes mainly involved in metabolism and basic cellular processes, such as adhesion, transport, and signaling (Fig. 12.3). On the other hand, the majority of genes regulated by the canonical ex-siRNAs of classes 1 and 2 correspond to sequences of unknown functions that are conserved in Mucorales and other fungi and show reduced expression during vegetative growth. However, it is worth noting that a number of those genes encode proteins involved in signal transduction and information storage and processing (Fig. 12.3), which may indicate the participation of these ex-siRNAs in the regulation of different and significant cellular processes.

Finally, class 4 corresponds to a tiny group of ex-siRNAs that requires Dcl-1, Ago-1 and the two RdRPs for their biogenesis (Table 12.1) (Nicolás et al. 2010; Cervantes et al. 2013). One of the exons included in this class encodes a conserved protein that colocalizes with other proteins in sites of polarized growth in yeast and other fungi. This, together with the fact that other exons code for proteins involved in mitochondrial metabolism and ribosomal function, may help to understand the phenotype of *dcl-1*$^{-}$ mutants, which show abnormal hyphal morphology and decreased growth rate (Nicolás et al. 2010).

Transcriptional Analysis

Whole-genome transcriptional analysis of wild-type and silencing mutant strains of *M. circinelloides* would provide an entire picture on the role of the silencing machinery in the regulation of endogenous functions, supplying a complete list of genes that are under the control of such machinery. Specific microarrays containing probes for all the predicted genes in the *M. circinelloides* genome were used to investigate differential mRNA accumulation in silencing mutants relative to the wild-type strain (our unpublished results). A total of 99 genes were differentially expressed in the *dcl-1*$^{-}$*/dcl-2*$^{-}$ mutant relative to the wild type during early exponential phase (p-adjust value <0.05). From these, 53 were upregulated in the mutant and 46 were downregulated. The majority of these genes showed the same pattern of expression in the *ago-1*$^{-}$ mutant, suggesting that they are directly or indirectly regulated by the canonical RNA silencing mechanism. However, it is noteworthy that the level of induction or repression relative to wild type was moderate, with a reduced number of genes showing a fold change higher than two.

The number of genes regulated by the silencing machinery dramatically increased when cells reached the stationary phase of growth. Under those conditions, a total of 1572 genes showed differential expression in the *dcl-1*$^{-}$*/dcl-2*$^{-}$ mutant relative to the wild type, half of which were upregulated and the other half downregulated (our unpublished results). Only a small proportion of the *M. circinelloides* genes differentially expressed in the *dcl-1*$^{-}$*/dcl-2*$^{-}$ mutant during the stationary phase were also differentially regulated in the *ago-1*$^{-}$ mutant (ca. 3 %), indicating that, besides regulating gene expression through the RNAi silencing mechanism, *M. circinelloides dicer* genes participate in noncanonical pathways to control gene expression. These noncanonical pathways are also participated by the RdRP-1 protein, since a relevant proportion of genes that are differentially regulated in the *dicer* double mutant, but not in *ago-1*$^{-}$, are similarly up- or downregulated in the *rdrp-1*$^{-}$ mutant. Again, gene expression changes provoked by the different mutations were moderate, indicating that the RNAi machinery is modulating gene expression rather than provoking drastic changes in it.

We assumed that at least some of the genes differentially expressed in the silencing mutants must be responsible for the phenotype shown by those mutants. In fact, preliminary analysis of the presumed functions of genes showing the highest differential expression in the *dcl*$^{-}$ and *ago*$^{-}$ mutants relative to wild-type (more than 3 fold change) identified genes that may be involved in growth, stress responses, and autophagy. Thus, several proteins that are downregulated up to 64 times in *dcl*$^{-}$ and *ago*$^{-}$ mutants are highly similar to cell wall modifying proteins that are crucial for cell wall integrity, septation and viability, and for adhesion and virulence (Willer et al. 2005; Munro et al. 2005). Particularly, one of them is highly similar to a *S. cerevisiae* G1/S-specific cyclin partner of the cyclin-dependent kinase PHO85, which is involved in the establishment or maintenance of cell polarity and in autophagy (Yang et al. 2010). Downregulation of these proteins could affect polar growth, asexual development, and stress responses in

M. circinelloides silencing mutants. Genes involved in establishment of polarity and septation are also differentially expressed in *T. atroviride dicer* mutants and it has been proposed that deregulation of these genes may be responsible for the phenotype shown by those mutants (Carreras-Villaseñor et al. 2013). Also proteins that are highly upregulated in the silencing mutants, as it should be expected from direct regulation through ex-siRNAs, have been identified. Most of them do not show similarity with proteins of known functions, although they contain domains presumably involved in gene regulation, which could explain the high number of genes deregulated in the silencing mutants. Interestingly, the most upregulated protein both in *dcl*$^{-}$ and *ago*$^{-}$ mutants corresponds to a class 2 ex-siRNA-producing locus. Thus, upregulation of mRNA correlates with downregulation of ex-siRNAs in the *dicer* and *ago* mutants. This protein shows a high similarity with the retrotransposon gag polyprotein of the Ty3/Gypsy family, which may confirm a direct role of the RNAi pathway in silencing retrotransposons during *M. circinelloides* vegetative growth. Most genes overlapping with other ex-siRNA loci were also deregulated in *dicer* and *ago* mutants, but the increase in mRNA accumulation in the mutant strains was modest, confirming that the RNAi pathway acts as a modulator of gene expression. However, it cannot be excluded the possibility that the *M. circinelloides* silencing mechanism may have higher effects on gene expression in different growth conditions or developmental stages.

Concluding Remarks

Fungi have contributed significantly to the understanding of the mechanisms underlying RNAi and its functions. Among basal fungi, studies with *M. circinelloides* have provided very valuable information, allowing a deeper understanding of the RNAi diversity among different organisms and highlighting the complexity of the silencing pathways in fungi. Analysis of the *M. circinelloides* ex-siRNAs and their target genes suggests that the silencing machinery could participate in very important biological processes in this fungus, such as light responses and pathogenesis (our unpublished results), thus expanding the range of endogenous sRNAs in eukaryotes and revealing a new role for them in fungi. Further studies may reveal exciting insights on the implications of these small RNAs molecules in a wide variety of functions, far beyond the initially proposed as a defense mechanism.

Acknowledgments This work was supported by the Spanish Ministerio de Ciencia e Innovación (BFU2009-07220, cofinanced by FEDER) to RMR-V and the Ministerio de Economía y Competitividad (BFU2012-32246, cofinanced by FEDER) to VG.

References

Alexander WG, Raju NB, Xiao H et al (2008) DCL-1 colocalizes with other components of the MSUD machinery and is required for efficient silencing in *Neurospora crassa*. Fungal Genet Biol 45:719–727

Calo S, Nicolás FE, Vila A et al (2012) Two distinct RNA-dependent RNA polymerases are required for initiation and amplification of RNA silencing in the basal fungus *Mucor circinelloides*. Mol Microbiol 83:379–394

Carreras-Villaseñor N, Esquivel-Naranjo EU, Villalobos-Escobedo JM et al (2013) The RNAi machinery regulates growth and development in the filamentous fungus *Trichoderma atroviride*. Mol Microbiol 89:96–112

Carthew RW, Sontheimer EJ (2009) Origins and mechanisms of miRNAs and RNAs. Cell 136:642–655

Catalanotto C, Azzalin G, Macino G et al (2002) Involvement of small RNAs and role of the *qde* genes in the gene silencing pathway in *Neurospora*. Genes Dev 16:790–795

Catalanotto C, Pallotta M, ReFalo P et al (2004) Redundancy of the two dicer genes in transgene-induced posttranscriptional gene silencing in *Neurospora crassa*. Mol Cell Biol 24:2536–2545

Cerutti H, Casas-Mollano JA (2006) On the origin and functions of RNA-mediated silencing: from protists to man. Curr Genet 50:81–99

Cervantes M, Vila A, Nicolás FE et al (2013) A single argonaute gene participates in exogenous and endogenous RNAi and controls cellular functions in the basal fungus *Mucor circinelloides*. PLoS ONE 8:e69283

Chang SS, Zhang Z, Liu Y (2012) RNA interference pathways in fungi: mechanisms and functions. Annu Rev Microbiol 66:305–323

De Haro JP, Calo S, Cervantes M et al (2009) A single dicer gene is required for efficient gene silencing associated with two classes of small antisense RNAs in *Mucor circinelloides*. Eukaryot Cell 10:1486–1497

Drinnenberg IA, Weinberg DE, Xie KT et al (2009) RNAi in budding yeast. Science 326:544–550

Emri T, Molnár Z, Szilágyi M et al (2008) Regulation of autolysis in *Aspergillus nidulans*. Appl Biochem Biotechnol 151:211–220

Fire A, Xu S, Montgomery MK et al (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 391:806–811

Ghildiyal M, Zamore PD (2009) Small silencing RNAs: an expanding universe. Nat Rev Genet 10:94–108

Grewal SI, Jia S (2007) Heterochromatin revisited. Nat Rev Genet 8:35–46

Grigoriev IV, Nordberg H, Shabalov I (2011) The genome portal of the Department of Energy Joint Genome Institute. Nucleic Acids Res 40:D26–D32

Hibbett DS, Binder M, Bischoff JF et al (2007) A higher-level phylogenetic classification of the Fungi. Mycol Res 111:509–547

Hoffmann K, Pawłowska J, Walther G et al (2013) The family structure of the *Mucorales*: a synoptic revision based on comprehensive multigene-genealogies. Persoonia 30:57–76

Jaskiewicz L, Filipowicz W (2008) Role of Dicer in posttranscriptional RNA silencing. Curr Top Microbiol Immunol 320:77–97

Jiang N, Yang Y, Janbon G et al (2012) Identification and functional demonstration of miRNAs in the fungus *Cryptococcus neoformans*. PLoS ONE 7:e52734

Kadotani N, Nakayashiki H, Tosa Y et al (2004) One of the two Dicer-like proteins in the filamentous fungi *Magnaporthe oryzae* genome is responsible for hairpin RNA-triggered RNA silencing and related small interfering RNA accumulation. J Biol Chem 279:44467–44474

Lau SKP, Chow W-N, Wong AYP et al (2013) Identification of microRNA-Like RNAs in mycelial and yeast phases of the thermal dimorphic fungus *Penicillium marneffei*. PLoS Negl Trop Dis 7:e2398

Lee SR, Collins K (2007) Physical and functional coupling of RNA-dependent RNA polymerase and Dicer in the biogenesis of endogenous siRNAs. Nat Struct Mol Biol 14:604–610

Lee HC, Chang SS, Choudhary S et al (2009) qiRNA is a new type of small interfering RNA induced by DNA damage. Nature 459:274–277

Lee HC, Li L, Gu W et al (2010) Diverse pathways generate microRNA-like RNAs and Dicer-independent small interfering RNAs in fungi. Mol Cell 38:803–814

Munro CA, Bates S, Buurman ET et al (2005) Mnt1p and Mnt2p of *Candida albicans* are partially redundant alpha-1,2-mannosyltransferases that participate in O-linked mannosylation and are required for adhesion and virulence. J Biol Chem 280:1051–1060

Nakayashiki H, Nguyen QB (2008) RNA interference: roles in fungal biology. Curr Opin Microbiol 11:494–502

Nakayashiki H, Kadotani N, Mayama S (2006) Evolution and diversification of RNA silencing proteins in fungi. J Mol Evol 63:127–135

Navarro E, Ruiz-Pérez VL, Torres-Martínez S (2000) Overexpression of the *crgA* gene abolishes light requirement for carotenoid biosynthesis in *Mucor circinelloides*. Eur J Biochem 267:800–807

Navarro E, Lorca-Pascual JM, Quiles-Rosillo MD et al (2001) A negative regulator of light inducible carotenogenesis in *Mucor circinelloides*. Mol Gen Genomics 266:463–470

Nicolás FE, Torres-Martínez S, Ruiz-Vázquez RM (2003) Two classes of small antisense RNAs in fungal RNA silencing triggered by non-integrative transgenes. EMBO J 22:3983–3991

Nicolás FE, de Haro JP, Torres-Martinez S et al (2007) Mutants defective in a *Mucor circinelloides dicer*-like gene are not compromised in siRNA silencing but display developmental defects. Fungal Genet Biol 44:504–516

Nicolás FE, Torres-Martínez S, Ruiz-Vázquez RM (2009) Transcriptional activation increases RNA silencing efficiency and stability in the fungus *Mucor circinelloides*. J Biotechnol 142:123–126

Nicolás FE, Moxon S, de Haro JP et al (2010) Endogenous short RNAs generated by Dicer 2 and RNA-dependent RNA polymerase 1 regulate mRNAs in the basal fungus *Mucor circinelloides*. Nucleic Acids Res 38:5535–5541

Nicolás FE, Torres-Martínez S, Ruiz-Vázquez RM (2013) Loss and retention of RNA interference in fungi and parasites. PLoS Pathog 9:e1003089

Ribes JA, Vanover-Sams CL, Baker DJ (2000) Zygomycetes in human disease. Clin Microbiol Rev 13:236–301

Romano N, Macino G (1992) Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences. Mol Microbiol 6:3343–3353

Ruiz-Vázquez RM, Torres-Martínez S (2003) Production of carotenes in Mucorales: structural and regulatory genes. In: Barredo JL (ed) Microorganisms for health care, foods and enzyme production. Research Signpost, Kerala, pp 69–84

Stajich JE, Harris T, Brunk BP et al (2012) FungiDB: an integrated functional genomics database for fungi. Nucleic Acids Res 40:D675–D681

Wei KF, Wu LJ, Chen J et al (2012) Structural evolution and functional diversification analyses of argonaute protein. J Cell Biochem 113:2576–2585

Willer T, Brandl M, Sipiczki M et al (2005) Protein O-mannosylation is crucial for cell wall integrity, septation and viability in fission yeast. Mol Microbiol 57:156–170

Xie Z, Johansen LK, Gustafson AM et al (2004) Genetic and functional diversification of small RNA pathways in plants. PLoS Biol 2:E104

Xue Z, Yuan H, Guo J et al (2012) Reconstitution of an Argonaute-dependent small RNA biogenesis pathway reveals a handover mechanism involving the RNA exosome and the exonuclease QIP. Mol Cell 46:299–310

Yamanaka S, Mehta S, Reyes-Turcu FE et al (2013) RNAi triggered by specialized machinery silences developmental genes and retrotransposons. Nature 493:557–561

Yang Z, Geng J, Yen W-L et al (2010) Positive or negative roles of different cyclin-dependent kinase Pho85-cyclin complexes orchestrate induction of autophagy in *Saccharomyces cerevisiae*. Mol Cell 38:250–264

Zhou J, Fu Y, Xie J et al (2012) Identification of microRNA-like RNAs in a plant pathogenic fungus *Sclerotinia sclerotiorum* by high-throughput sequencing. Mol Genet Genomics 287:275–282

Chapter 13
Regulation of Pericentric Heterochromatin by ncRNA in *Schizosaccharomyces pombe*

Mikel Zaratiegui

Abstract Heterochromatin is a conserved chromatin arrangement with roles in transcriptional regulation and genome integrity. It is characterized by regular nucleosomal arrays decorated with specific histone post-translational modifications. These recruit homologs of the HP1 protein that mediate the formation of a tightly compacted chromatin structure refractory to virtually all nucleic acid transactions. Heterochromatin fibers organize into higher order structures that shape a large part of subnuclear organization. Heterochromatin coats a diverse set of genetic elements whose only common feature is their repetitive nature. These include transposable elements (TE), satellite repeats, and subtelomeric repeats. The mechanisms that target these elements for compaction into heterochromatin have been elusive for a long time, with tantalizing indications of a central role for RNA in this process. In the last decade, work in model organisms from fungi to plants and metazoans have revealed the highly dynamic nature of heterochromatin, and unexpected relationships with fundamental cellular processes, guided by non-coding RNA. The deepest mechanistic insights into this phenomenon have come from the study of heterochromatin in the fission yeast *Schizosaccharomyces pombe*. In particular, the heterochromatin that covers pericentric repeats has been intensely scrutinized, because of the tandem repeat structure reminiscent of centromeres of higher organisms. This chapter summarizes the current view of the organization and regulation of pericentric heterochromatin in *S. pombe* with a focus on involvement of non-coding RNA.

M. Zaratiegui (✉)
Department of Molecular Biology and Biochemistry, Rutgers, The State University of New Jersey, Piscataway, NJ, USA
e-mail: zaratiegui@biology.rutgers.edu

A. Sesma and T. von der Haar (eds.), *Fungal RNA Biology*,

DOI: 10.1007/978-3-319-05687-6_13,

Contents

Introduction

Heterochromatin: The "Dark Matter" of the Genome

Chromatin is the combination of DNA and the protein complement that regulates it. The main components of chromatin are globular complexes called nucleosomes that package the DNA and serve as a smart storage system, marking the coated sequences with regulatory signals. These signals mediate the adoption of local conformations that determine the accessibility of the DNA to the transcription and recombination machineries. Depending on the degree of compaction of chromatin we can distinguish two forms of organization, euchromatin, and heterochromatin. While euchromatin is open and accessible, heterochromatin is a specialized form of chromatin with a tightly packed structure. Heterochromatin can be facultative, which appears only in certain circumstances as dictated by the regulatory or developmental needs of the cell. But much of heterochromatin is constitutive, and the regions it covers are always in a closed conformation. The influence of the compacted structure of constitutive heterochromatin in its underlying elements is obvious in the inhibition of meiotic recombination, yielding genetic distances that

are disproportionate to physical distances. Additionally, heterochromatin was soon observed to be transcriptionally inert. This can be explained by the structure of heterochromatin preventing binding of transcription factors and productive transcription. Notably, it was observed that heterochromatin can spread *in cis*, silencing nearby genes. This phenomenon, termed Position Effect Variegation (PEV) (Allshire et al. 1994), was the first indication that not all loci in the genome are equal for gene expression. Due to the lack of transcriptional activity, which was taken to represent a low density of genes and other interesting elements, heterochromatin has received much less attention than its more gene rich cousin, euchromatin. Nevertheless, its importance as a nuclear and chromosomal organizer, and as a protector of genome integrity, was recognized since very early on (Yunis and Yasmineh 1971).

Functions of Constitutive Heterochromatin

Constitutive heterochromatin covers regions of the genome that are highly repetitive, and its high compaction prevents transcription as well as recombination of the repeat elements. These are important functions because many repetitive sequences are derived from Transposable Elements (TE), potentially harmful genetic elements capable of moving and replicating within the genome. These selfish parasitic sequences must be silenced to avoid their rampant spread and the mutation and genomic instability that they can cause.

The other main types of sequences coated by constitutive heterochromatin are highly repetitive arrays of non-coding elements called satellite DNA. Satellite repeat monomers are usually very short (around the length of a nucleosome), but elements of a few Kb in length are not uncommon. Regardless of the size of their constitutive monomers, satellite arrays of multimers arranged in tandem head-to-tail fashion can expand to enormous multi-megabase sizes, with a large degree of variability in size observed between strains in *Neurospora* (Smith et al. 2012) and *Schizosaccharomyces pombe* (Steiner et al. 1993). Satellite repeats lack sequence conservation across species but are homogeneous within a given genome. All of these features have led to a model for satellite repeat amplification and homogenization that involves gene conversion and unequal crossover.

Satellite repeats are non-coding, and their transcription may have toxic effects. As such, their transcriptional silencing might be an important role of heterochromatin. However, the main role of heterochromatin formation in satellite repeats appears to be in chromosome segregation. Mutants that affect satellite repeat heterochromatin exhibit segregation problems like lagging chromosomes in late anaphase that result in rearrangements and aneuploidies (Ekwall et al. 1999; Peters et al. 2001; Volpe et al. 2003), due to a failure of centromere function. The relationship between satellite repeats and centromeres is illustrated by the fact that the kinetochore organizing regions of centromeres are almost invariably associated with, and in many cases even composed of, satellite repeats. Moreover, in any given

genome, the most abundant satellite repeat is usually located in centromeric regions (Melters et al. 2013). The current understanding of the role of heterochromatin in centromere function invokes a structural role, providing rigidity through higher order arrangements allowing the precise tension forces that determine the chromosome alignment required for equal segregation. Furthermore, HP1 is thought to serve a role in retention of cohesin at centromeres in prophase to prevent their premature separation (Bernard et al. 2001). Consistently, cohesin and heterochromatin mutants exhibit synthetic lethality in fission yeast.

Satellite repeats are also commonly interspersed with TE, in particular retrotransposons, which sometimes exhibit a preference for insertion into heterochromatic regions. In some cases, like *Neurospora* (Cambareri et al. 1998), *Aspergillus nidulans* (Nielsen et al. 2001) and *Schizosaccharomyces japonicus* (Rhind et al. 2011), it appears that the genome has organized the centromeres using heterochromatic mutated remnants of TE. Conversely, some satellite repeats may have a transposon origin. A domesticated *Pogo* family transposase, CENP-B, is probably the only DNA binding protein with a conserved binding site frequently found in centromeric satellites. CENP-B binds to satellite repeats through its conserved CENP-B box sequence, where it regulates heterochromatin and centromere formation (Kipling and Warburton 1997). Notably, in *S. pombe* CENP-B is instead found at the long terminal repeats of a TE unrelated to *Pogo*, where it silences transcription through recruitment of heterochromatin factors (Cam et al. 2008). This almost mystical relationship between TE and centromeres has puzzled researchers for decades, and may suggest that TE silencing and the heterochromatin component of centromeres arise from similar structures.

Molecular Determinants of Constitutive Heterochromatin

Like other types of chromatin, heterochromatin exhibits characteristic molecular features that determine its physical and genetic properties. These features have been investigated through genetic screens looking for suppressors of PEV (SuVar). At its basis, the histone code that decorates heterochromatic nucleosomes shows hypoacetylation of Histone 4 (H4) and the N terminus of Histone 3 (H3) as well as dimethylation and trimethylation of Lysine 9 of H3 (henceforth collectively referred to as H3K9me). These marks are deposited by the concerted actions of Histone Deacetylases (HDAC) and the SET domain histone methyl transferase family SuVar3-9. Such a combination of modifications is read by the chromodomain of HP1, which binds to H3K9me in a manner that can be regulated by phosphorylation and acetylation of nearby residues in H3 (Yamada et al. 2005; Fischle et al. 2005). HP1 can recruit SuVar3-9 to methylate H3K9 in adjacent nucleosomes, providing a new docking platform for additional HP1 and a mechanism for the spreading *in cis* that causes PEV.

In many organisms, heterochromatin is also associated with DNA methylation (DNAme). It has been difficult to establish the order of deposition of the two

chromatin marks, but it appears that H3K9me can direct DNAme in plants and humans (Lehnertz et al. 2003; Du et al. 2012). This relationship is conserved in fungi that feature DNAme: in *Neurospora*, DNA methylation requires the activity of the SuVar3-9 family member DIM-5 and HP1, which in turn recruits the DNA methyltransferase DIM-2 (Honda et al. 2012).

While the molecular features of constitutive heterochromatin are well understood, little is known about the mechanisms that recruit the chromatin modifiers to deposit heterochromatic marks on repetitive DNA. The problem is compounded by the lack of sequence conservation of heterochromatic DNA, which indicates that no single universal DNA binding specificity is responsible for the initial mark deposition. Some sequence specificity is nevertheless necessary to prevent the formation of ectopic heterochromatin. The study of pericentric heterochromatin of fission yeast has revealed the strong implication of non-coding RNA (ncRNA) in this specificity.

The Pericentric Heterochromatin of *S. pombe*

Fission yeast centromeres have received considerable attention because they exhibit extensive similarities with the centromeres of higher organisms. Unlike *Saccharomyces cerevisiae*'s point centromeres, *S. pombe* centromeres are epigenetically determined, show extensive repetitive structure, and require the participation of constitutive heterochromatin.

Genetic Structure of S. pombe Centromeres

The genetic structure of the three fission yeast centromeres was established through chromosome walking from centromere-linked genes (Nakaseko et al. 1986, 1987; Clarke et al. 1986). This analysis revealed that repeats of two sequences, named dg (abbreviation of "dogentai," Japanese for kinetochore) and dh (named thus because it was identified after dg), constituted the bulk of the centromeric DNA (Fig. 13.1).

Parts of both dg and dh are present in all centromeres (Figs. 13.1 and 13.2), but the repetitive arrangements that they form are specific to each of them (Fishel et al. 1988). These arrangements are repeated in tandem conformation, and form the outer repeats (otr). More detailed analysis revealed the presence of additional repeats, named inner most repeats (imr), separating dg and dh from a central domain of unique sequence(Clarke and Baum 1990). Together, imr and otr form two large domains in inverted orientation that flank the central core in each centromere. Depending on the number of repeat units on either side of the central core, the pericentric repeats may form symmetric or asymmetric structures around it.

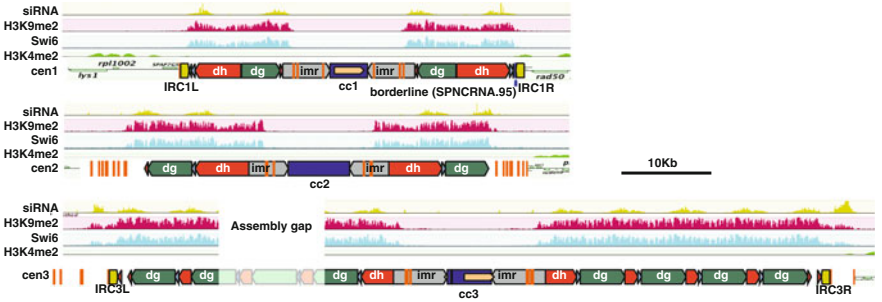


Fig. 13.1 Genetic and epigenetic structure of *S. pombe* centromeres. The centromeres from chromosomes 1, 2, and 3 are depicted to scale. Vertical yellow lines depict tRNA genes. Central core (cc) regions in dark. H3K9me2 and *Swi6* ChIP-chip data are from (Cam et al. 2005), siRNA data are from (Halic and Moazed 2010). The size of the unassembled gap is estimated from the pulsed-field electrophoresis measuring of centromere size. Repeat structure presumed to constitute the assembly gap is depicted as grayed out areas

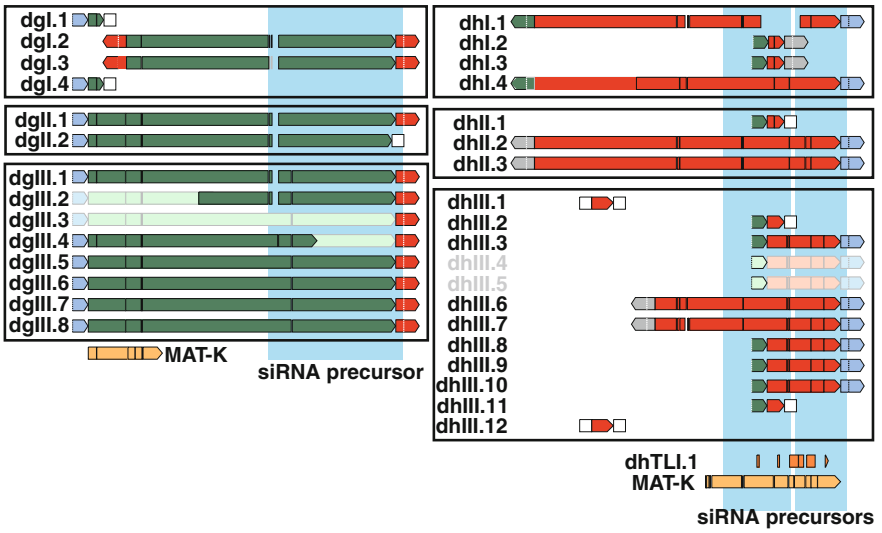


Fig. 13.2 Alignment of dg and dh repeats. imr repeats in *gray*. siRNA precursors are highlighted. Regions of homology with CenH/MAT-K and dh sequence found in a subtelomeric gene (dhTLI.1) are depicted under the alignment. Data courtesy of Derek Bartlem

Throughout this work we will refer to the structure found in the type strain 972, which was the subject of the genome sequencing effort (Wood et al. 2002), but it is pertinent to comment on the observed variability. There is significant polymorphism in the size of centromeric DNA between laboratory and wild strains arising from differences in the number of repeats (Steiner et al. 1993). The otr can even whittle down to one single repeat on either side of the central core. However, the

repeat units found within each centromere appear to be constant (Fig. 13.2). A 1.7 kb region within the dg repeat is strikingly well conserved (97 %), and deletion of this region impairs centromeric function (Baum et al. 1994). The central domains can also invert with respect to the rest of the chromosome (Steiner et al. 1993). These observations indicate that there is an active interaction between the repeats in each centromere that normalizes their sequence and leads to gains and losses in copy number. Meiotic recombination in the centromere is strongly suppressed (Clarke et al. 1986), so this activity could be the result of mitotic Homologous Recombination (HR) within each chromosome.

Ab initio sequence analysis found no protein coding genes in the pericentric repeats or the central core. Consistently, no polyadenylated transcripts were detectable using the repeat sequence as a probe (Fishel et al. 1988). However, clusters of tRNA genes are detectable in the imr fragments that separate the otr from the central core and from the unique sequence outside the centromere (Kuhn et al. 1991). It was noted that these clusters coincided with a transition in the chromatin structure of the central and repeated domains, pointing at a role in centromeric function.

It appears from the variable centromeres found in the wild that the structure of a minimal functional centromere is two inverted repeat dg/dh monomers around a central core (Steiner et al. 1993). When tested in circular minichromosomes, the combination of the central core of centromere 2 and portions of the dg repeat is sufficient to confer appreciable segregation (Baum et al. 1994; Marschall and Clarke 1995; Folco et al. 2008). Additional copies of dg increased the functionality of this minimal centromere. It is clear from these experiments that the central core and the otr provide separate functions, both necessary but neither sufficient, for centromere function.

Chromatin Structure

Throughout most eukaryotes the preferred genetic substrate for kinetochore assembly are repetitive sequences. However, these repeats are often dispensable, as illustrated by neocentromeres that can form without their participation, and the repeats alone are seldom sufficient to assemble a functional centromere. There is therefore a strong epigenetic component to centromere specification. The clear distinction between the different domains that constitute the *S. pombe* centromere has allowed us to make connections between their chromatin structure and centromeric function.

Micrococcal nuclease digestion revealed a stark contrast in the chromatin organization of the central and pericentric domains. While the pericentric repeats showed highly regular arrays of nucleosomes, the central core had an atypical pattern of diffuse digestion (Polizzi and Clarke 1991). The transition occurs at the tRNA gene clusters that separate the two regions (Kuhn et al. 1991). The central domain and the inner part of imr was later shown to be composed of nucleosomes that substitute H3 with its variant CENP-A (Takahashi et al. 2000), as well as some

canonical nucleosomes decorated with the euchromatic mark H3K4me2 (Cam et al. 2005). CENP-A nucleosomes adopt a more compact conformation, possibly contributing to the stark difference in the chromatin structure. CENP-A is a universally conserved H3 variant that serves as the cornerstone of kinetochore assembly. The mechanisms that deposit CENP-A in the central core are still not well understood, but they appear to employ multiple collaborating pathways (reviewed in (Quénet and Dalal 2012)).

The otr region exhibits heterochromatic properties, such as suppression of meiotic recombination and lack of detectable transcriptional activity (Fishel et al. 1988). Importantly, transgenes inserted into the pericentric repeats are silenced in a variegated manner (Fishel et al. 1988; Allshire et al. 1994) reminiscent of the classic PEV phenomenon first described in *Drosophila*. The degree of silencing of the transgene depends on the exact position of the insertion, with some locations showing stronger silencing than others. Insertions into the central core are also partially silenced (Allshire et al. 1994). Using the PEV system in forward genetic screens it has become clear that silencing of these two regions uses distinct mechanisms (Ekwall et al. 1999; Javerzat et al. 1999). The cause of the heterochromatic structure of the pericentric repeats can be traced to *cis*-acting sequences within them. It is possible to form heterochromatin domains with fragments of the dg and dh repeats inserted ectopically (Partridge et al. 2002). Importantly, a fragment with almost complete homology to pericentric repeats (cenH) acts as a heterochromatin organizer in the mating type locus (Ayoub et al. 1999; Hall et al. 2002), where it is necessary for efficient mating type switching (Grewal and Klar 1997; Jia et al. 2004b).

A genome-wide study using Chromatin Immunoprecipitation coupled to microarray analysis (ChIP–chip) provided a detailed view of the organization of centromeric chromatin (Cam et al. 2005). This study confirmed that the entirety of the otr region as well as the outermost side of imr were covered with H3K9me2 and me3 nucleosomes, as well as the HP1 homolog *Swi6* that binds to it (Fig. 13.1). Heterochromatin dropped sharply at the transition zones determined by the tRNA gene clusters. Outside of the otr, and separated from them either by clusters of tRNA genes or other repeats, lie domains of euchromatin with its characteristic H3K4me mark. The highly compartmentalized structure suggested the activity of strong boundary elements between the different domains. This activity is provided by the tRNA genes (Scott et al. 2006) in the boundaries that have them, but other elements must also act as barriers in those that don't (Cam et al. 2005; Keller et al. 2013).

The molecular distinction between the central and pericentric domains is mirrored in their subnuclear localization in interphase (Kniola et al. 2001): The CENP-A domain is anchored to the nuclear envelope, close to the spindle pole body, with the *Swi6* domains closely associated and forming cytological heterochromatin domains. This layered structure can also be observed in human centromeres.

In summary, the chromatin structure of fission yeast centromeres is composed of two large heterochromatic domains flanking a central region of CENP-A

chromatin that will recruit the kinetochore. This arrangement is also present in the centromeres of other organisms, and highlights both the fundamentally epigenetic nature of centromeres and the function of constitutive heterochromatin as a helper of centromere function.

Mechanisms of Pericentric Heterochromatin Deposition

The phenomenon of PEV in pericentric repeats offered the possibility of performing forward genetic screens analogous to those in *Drosophila* (Ekwall et al. 1999). These studies were complemented by efforts to dissect the process of mating type switching, which also partially depends on proper heterochromatin function (Ekwall and Ruusala 1994; Thon et al. 1994). More recently, the picture has been completed with high-throughput surveys of genetic interaction that revealed novel heterochromatin factors (Roguev et al. 2008), and with biochemical characterization of the protein complexes they form. The factors identified in these approaches permitted the characterization of the molecular underpinnings of heterochromatin.

The H3K9me/HP1 Axis

The *swi6* gene, a homolog of HP1, was identified in a mating type switching screen (Klar and Bonaduce 1991). Its cloning revealed homology to chromodomain (chromatin modifier domain) factors HP1 and Polycomb (Lorentz et al. 1994). Localization at centromeres established it as the HP1 homolog in fission yeast (Ekwall et al. 1995), and it was later shown to bind H3K9me2 (Bannister et al. 2001). *Swi6* can multimerize. Cooperative binding of *Swi6* to arrays of H3K9me nucleosomes, and their subsequent multimerization determines the highly condensed structure that is characteristic of heterochromatin (Canzio et al. 2011; Al-Sady et al. 2013).

The complementation groups that showed consistently strongest loss of silencing in both centromere and mating type locus were named *clr4* (iodine-clear colonies) and *rik1* (recombination in K). *Clr4* was identified as the H3K9me transferase by virtue of its homology with SET (SuVar39, Enhancer of Zeste, Trithorax) domain factors (Rea et al. 2000; Nakayama et al. 2001). *Rik1*p is a paralog of the DNA repair factor *Ddb1*, and forms a complex with two other factors that are required for HP1 localization, *Raf1*/*Dos1* and *Raf2*/*Dos2* (*Rik1* Associated Factor/Delocalization Of *Swi6*), and with the E3 Ubiquitin ligase Pcu4 (Horn et al. 2005; Jia et al. 2005; Thon et al. 2005; Li et al. 2005). This complex is highly reminiscent of the *Ddb1* system that serves in selection of ubiquitination substrates (Buscaino et al. 2013). While the ubiquitination target is unknown, the activity is necessary for H3K9me deposition. The *Rik1* complex interacts with

Clr4 to constitute the CLRC (*Clr4-Rik1-Cul4*) complex (Sadaie et al. 2004; Hong et al. 2005), suggesting that the two activities are coupled. Tethering the CLRC complex to chromatin, via fusion with sequence-specific DNA binding domains, is sufficient to direct H3K9me deposition and *Swi6* binding to ectopic sites. This synthetic heterochromatin is sufficient to confer centromeric activity to adjacent sequences from the central core (Folco et al. 2008; Kagansky et al. 2009).

Clr4 also has a chromodomain, as do two other factors necessary for proper heterochromatin formation, *Chp1* and *Chp2*. The binding of these chromodomain factors to the available H3K9me depends on the presence of other H3 tail modifications like H3K14 acetylation and H3S10 phosphorylation, adding to the complexity of regulation that is layered on top of H3K9me (Sadaie et al. 2004; Fischle et al. 2005; Xhemalce and Kouzarides 2010; Alper et al. 2013).

The major determinant of heterochromatin localization is therefore the recruitment of the *Clr4* containing complex and the subsequent deposition of H3K9me. Weaker loss of silencing mutants identified the pathways that collaborate to prepare chromatin for heterochromatin assembly and localize CLRC. These can be summarized in two nonoverlapping pathways: the SHREC chromatin remodeler and RNA interference (RNAi). Mutations in either pathway lead to a partial reduction of H3K9me, but double mutants affecting both show complete elimination of this mark (Yamada et al. 2005; Marina et al. 2013).

SHREC

SHREC (*Snf2*/HDAC Containing Repressor Complex) is a complex of 4 core proteins that include the HDAC *Clr3* and the ATP-dependent chromatin remodeler *Mit1* (Sugiyama et al. 2007). *Clr3* deacetylates H3K14, and prevents aberrant H3S10 phosphorylation and RNA polymerase II (RNA pol II) recruitment (Yamada et al. 2005). Thus, SHREC prevents deposition of euchromatic marks and facilitates binding of *Swi6*. SHREC is recruited to the mating type locus heterochromatin by the CREB transcription factor *Atf1* (Jia et al. 2004a; Yamada et al. 2005), but it can spread to nearby regions with the participation of *Swi6* (Sugiyama et al. 2005) and/or the chromodomain protein *Chp2* (Motamedi et al. 2008; Fischer et al. 2009) to form large heterochromatin domains. The localization of SHREC to pericentric repeats requires an RNA binding factor (Marina et al. 2013), *Seb1*, that probably binds to the same ncRNA that RNAi processes (described below). In summary, SHREC can be recruited to heterochromatin nucleation centers by several mechanisms, from where it spreads across large regions to remodel chromatin and facilitate heterochromatin deposition by CLRC and *Swi6*.

The RNAi Pathway

RNAi was discovered as a cytoplasmic system of sequence-specific post-transcriptional silencing, with TE as major targets (see also Chap. 11). Subject of intense research, the last 15 years have yielded a vast amount of information on this phenomenon and its critical roles in gene regulation. RNAi is the collective name for a set of very diverse pathways, specializing on regulation of different genic and non-genic targets. Their common defining feature is the use of short stretches of RNA between 20 and 30 nucleotides in size as sequence specific determinants of their targets. The small RNAs, or short interfering RNA (siRNA) are usually the product of processing of larger precursors through specialized nucleases. The most usual source of siRNA is cleavage of double-stranded RNA (dsRNA) by the bidentate RnaseIII Dicer , but other nucleases can process single-stranded precursors (as is the case for piRNA), or they can be the direct products of RNA-dependent RNA polymerases. Regardless of their source, the siRNA, bound to a member of the Argonaute/PIWI family of RNAses, constitutes the effector complexes that are directed by base-pairing of the siRNA to target RNA.

Mutants of RNAi pathways in plants and metazoans often have developmental defects, usually as a result of loss of microRNA (miRNA). These are siRNA derived from hairpin precursors that are directed against protein coding genes and provide fine-tuning regulation of their expression. Fungal miRNA have proved elusive, but there are indications that *Neurospora* may have miRNA-like pathways (Lee et al. 2010) (see Chap. 11). Besides miRNAs, other RNAi pathways with multiple types of targets, from viruses and TE to protein coding genes, are at work in virtually all organisms. The glaring exception is *S. cerevisiae*, which seemingly lost RNAi as an adaptation to the dsRNA killer virus (Drinnenberg et al. 2011). This wide diversification and specialization of RNAi pathways in plants and metazoans, and their multiple interactions and redundancies, made it difficult to learn the fundamental characteristics of RNAi. This directed attention to *S. pombe* as a fungal model organism with a minimal RNAi pathway.

S. pombe only has one homolog of each of the three main types of factors driving RNAi. One Dicer, *Drc1*; one Argonaute/PIWI factor of the Argonute clade, *Ago1*; and one RNA-dependent RNA polymerase, *Rdp1*. A reverse genetic approach to tease out the function of RNAi in *S. pombe* revealed that mutants in these factors exhibited defective heterochromatin formation (Volpe et al. 2002; Hall et al. 2002), with decreased H3K9me and *Swi6* binding. *ago1* was also discovered to be allelic to one of the mutants isolated in the PEV forward genetic screens, *csp9* (Ekwall et al. 1999; Volpe et al. 2002). RNAi mutants were later shown to display mitotic and meiotic chromosome segregation defects common to known heterochromatin factors (Volpe et al. 2003; Hall et al. 2003).

Importantly, RNAi factors (*Rdp1* (Volpe et al. 2002), *Ago1* (Cam et al. 2005) and *Dcr1* (Woolcock et al. 2011)) localize to the pericentric repeats. This indicates that their role in heterochromatin deposition is direct, rather than regulating

expression of other heterochromatin factors. Since the natural substrate of RNAi is RNA, it would follow that the target of RNAi in *S. pombe* are transcripts arising from the supposedly silent pericentric repeats. Cloning and sequencing of *S. pombe* siRNA revealed that the siRNA indeed corresponded to the pericentric repeats (Reinhart and Bartel 2002; Verdel et al. 2004). These discoveries suggested the exciting possibility that siRNA could be sequence-specific determinants of heterochromatin formation. This would bypass the puzzling lack of conservation in heterochromatic DNA that made the universal participation of sequence specific DNA binding factors unlikely: repetitive DNA could provide their own sequence specificity by expressing RNA that would be processed and used by RNAi.

This phenomenon was also the first example of a nuclear role for RNAi. Fission yeast *Dcr1* localizes to the nuclear periphery, associated with the nuclear pore complex (Emmerth et al. 2010), in contrast with human DICER1 which is exclusively cytoplasmic. Nuclear retention of *Dcr1* is regulated by stress signals, implicating RNAi in the response to stress (Woolcock et al. 2012). Interestingly, the zinc finger domain required for nuclear retention is conserved in pathogenic yeasts (Barraud et al. 2011). Being specific to fungi this mechanism of nuclear localization could constitute a potentially druggable target.

The concentration of RNAi activity in the nucleus does not prevent cytoplasmic effects. *Ago1*, *Dcr1* and *Rdp1* can be detected in cytoplasmic puncta (Carmichael et al. 2006). Reprogramming of RNAi via hairpins with sequence directed against a transgene have mostly a post-transcriptional effect, with no detectable deposition of H3K9me or loss of transcriptional activity (Sigova et al. 2004), although transcriptional repression can be observed in some cases (Simmer et al. 2010). Two *Ago1* interactors, *Arb1* and *Arb2*, implicated in maturation of the functional siRNA-*Ago1* nucleoprotein are also cytoplasmic (Buker et al. 2007), opening the possibility of a shuttling *Ago1*. A shuttling *Ago1* would lose the information of point of origin as it exits the nucleus, enabling silencing to homologous sequences *in trans* once it is reimported.

RNAi Processing of Centromeric ncRNA

The Precursor ncRNA

In order for RNAi to generate siRNA, there must be at least some level of transcription coming from the pericentric repeats. Since in the wild type case any transcription product is rapidly turned over by the RNAi machinery, little to no RNA is detectable by classic techniques like Northern Blot or RT-PCR. In the case of heterochromatin and RNAi mutants discrete bands become visible due to the simultaneous loss of transcriptional and post-transcriptional silencing. The transcription signal is strongest in *Clr4* and *Rik1* mutants. Mutants in RNAi genes show

milder upregulation, and SHREC mutants very slight upregulation (Zaratiegui et al. 2011). However, double mutants in RNAi and *Clr3* show very strong signals, exceeding that seen in *Clr4*/*Rik1* mutants. This correlates with the heterochromatic marks present: double RNAi/*Clr3* mutants show a complete loss of H3K9me2. The RNAi machinery and SHREC have therefore a partial division of labor in the transcriptional silencing of the pericentric repeats. Interestingly, siRNA levels are elevated in *Clr3* mutants, indicating that the still functional RNAi pathway is processing the transcripts resulting from loss of SHREC-mediated transcriptional repression (Sugiyama et al. 2007).

The bands detectable by Northern blot are suggestive of discrete transcriptional units of variable length. Also, probes directed to some regions of the pericentric repeats show no detectable signal even in strong mutants with complete loss of silencing. Deep sequencing of siRNA exhibited discrete regions of siRNA accumulation of both forward and reverse orientation, separated by regions with no detectable RNAi activity coinciding with the untranscribed regions (Cam et al. 2005; Zaratiegui et al. 2011), reflecting the extension of the siRNA precursors. Transcription and siRNA generation are therefore not pervasive to the whole heterochromatin region and are instead restricted to particular loci within otr and imr.

Cloning and sequencing of the siRNA precursors from *swi6* and *dcr1* mutants revealed their structure. The precursors are derived from loci that coincide with the highest abundance of siRNA, and range in size from 1 to 6 kb. Curiously, the precursor loci are transcribed in both forward and reverse orientations, with the transcription initiation region for one strand roughly colocalizing with the termination region of the complementary strand. One of the orientations is usually more intensely transcribed than its reverse partner, possibly reflecting different efficiencies in transcription initiation from their corresponding promoters. The precursors are often polyadenylated and are enriched in the poly(A) + fraction. This is a strong indication that the responsible polymerase is RNA polymerase II. Introns are sometimes observable (Chinen et al. 2010), but are never spliced out with complete efficiency. Both the transcription initiation sites and the cleavage and polyadenylation sites are highly variable, contributing to the size heterogeneity (Zaratiegui et al. 2011). Indeed, low levels of siRNA are detectable well beyond the initiation/termination regions. The weak efficiency of the cleavage and polyadenylation sequences contributes to RNAi processing of the precursors, probably because the nascent RNA is retained at the site of transcription to serve as an anchoring point (Yu et al. 2013). The efficiency of transcription initiation, splicing and polyadenylation are probably influenced by the chromatin environment, but the lack of information on the precursors as they are expressed in wild type cells makes it difficult to know the exact processing preferences.

Since the transcripts initiate and terminate in specific regions within the pericentric repeats, it is the different arrangements of repeats in each of the centromeres that yields the observed size heterogeneity (Fig. 13.3). While the dh repeat has three initiation and termination regions in its sequence, both the imr and dg repeats only have one (Fig. 13.2). Thus, a precursor can be completely self-contained

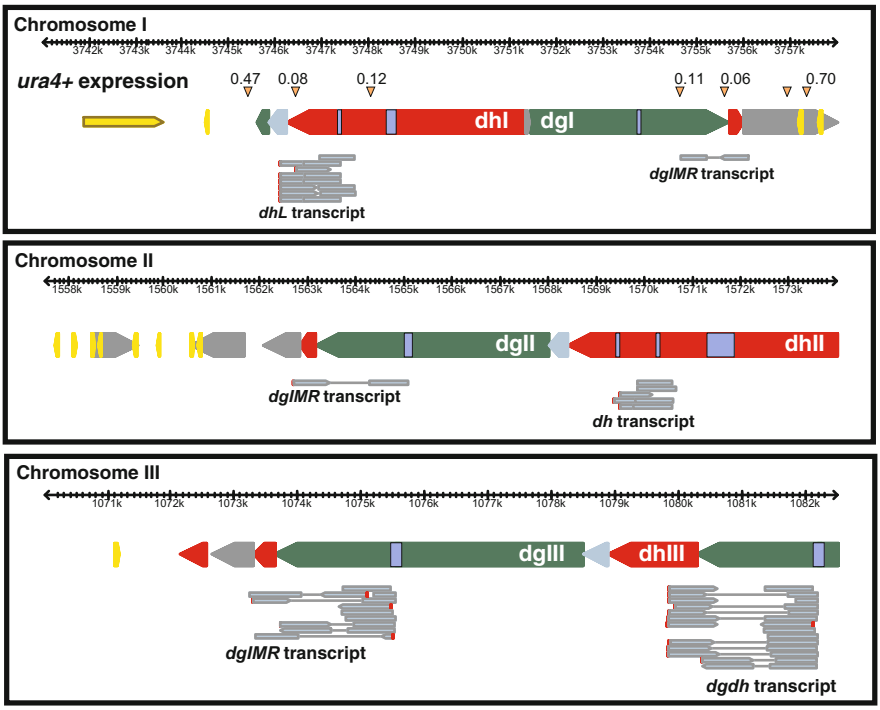


Fig. 13.3 The main siRNA precursors in their chromosomal context. Sequence from cDNA depicted as light rectangles. *ura4+* expression: Arrowheads mark the positions of *ura4+* reporter insertions, and the number above depicts the expression as a fraction of the euchromatic expression (data from (Allshire et al. 1994)). Poly(A) tails, when detected, are depicted as red ends of the precursors. Gaps in the sequence are shown as thin lines connecting the blue rectangles. Ruler depicts coordinates in May 2002 assembly (Wood et al. 2002). Mapping data courtesy of Derek Bartlem

within a repeat type (the dh transcript) or initiate in the promoter present in the imr and end in the dg repeat or vice versa (the imrdg transcript, present in all chromosomes). Combinations of initiation and termination regions constitute the repertoire of siRNA precursors.

Interesting patterns emerge when we cross-reference the localization of the precursors with data about the chromatin substructure of the pericentric repeats. The most striking is perhaps the fact that the precursors coincide with the localization of transgene insertions that are most efficiently silenced in models of PEV (Allshire et al. 1994; Irvine et al. 2006) (Fig. 13.3). The insertions become part of the precursor, forming a co-transcript and yielding siRNA albeit at lower levels than the pericentric sequences (Irvine et al. 2006). Interestingly, in RNAi mutants the pericentric repeats retain some H3K9me, but the transgene insertions become completely euchromatic (Sadaie et al. 2004). The remaining H3K9me in the repeats depends on *Clr3*, but mutations in *Clr3* do not lead to a decrease in

H3K9me in the insertions. This suggests that, while the pericentric repeats are targeted by SHREC via sequence-specific signals, insertions of non-centromeric repeat sequences into siRNA precursors become the exclusive purview of RNAi, which is solely responsible for spreading H3K9me into them.

Interestingly, a minimal sequence capable of recruiting ectopic heterochromatin, L5, contains the dg promoter (Partridge et al. 2002). The dg promoter is also contained within the highly conserved region necessary for centromeric function in circular minichromosomes (Baum et al. 1994). The promoters present in dh and imr repeats coincide with regions that lose nucleosome occupancy in heterochromatin mutants (Garcia et al. 2010). In this respect siRNA precursor loci behave much like regular RNA pol II-transcribed genes, in that when active they show a nucleosome-free region (NFR) upstream of the transcription start site. Interestingly, the requirements for keeping these NFR occluded with nucleosomes and therefore transcriptionally silent are quite variable. While all NFR open up in *clr4* mutants, some of them require the combinatorial activity of components of the SHREC complex *Clr3* and *Mit1*, and the chromodomain protein *Chp2* with others independent of *Clr3*. This is probably reflected in the sensitivity of the different promoters to *Clr3* mutations as revealed by the accumulation of extra siRNA in dg but not in dh sequence. This differential regulation is also observed in transcription run-on experiments of the dh transcript (Volpe et al. 2002): one of the strands (forward) is silent while the other (reverse) is active even in the wild type strain. Strand-specific RT-PCR of this same locus shows upregulation of the forward strand in *swi6* mutants, but not of the reverse. Therefore, only the forward strand is transcriptionally silenced by heterochromatin, while the reverse strand is capable of transcribing and is processed post-transcriptionally by RNAi. The assembly of the transcription machinery on these diverse promoters hasn't been investigated, but their importance is highlighted by the fact that they constitute portable heterochromatin nucleation sequences.

As suggested by the detectable polyadenylation and splicing of the siRNA precursors, the polymerase responsible for their transcription is RNA pol II. Alleles in several subunits of the RNA pol II holoenzyme affect PEV without grossly disrupting normal mRNA transcription (Kato et al. 2005; Djupedal 2005). Interestingly, an allele of subunit *rpb7* affected RNAi processing and heterochromatin by losing transcription of only the reverse strand in the dg-imr transcript in centromere 1 (Djupedal 2005), once again showing the differential sensitivity of siRNA precursor promoters to heterochromatin mutations. In summary, transcription of siRNA precursors is a highly regulated process, but the difficulty of separating post-transcriptional and transcriptional effects makes it difficult to investigate the regulatory activities responsible.

RNAi Processing: The Self-reinforcing Loop

The mechanisms behind siRNA generation and its coupling to heterochromatin silencing have been dissected by the collective effort of many groups, and are still yielding surprises even over 10 years after the initial discovery. The first such surprise was that, just like H3K9me depends on RNAi, RNAi depends on H3K9me: mutation of *clr4* abrogates siRNA generation. Therefore, RNAi cannot be definitely put in a clear pathway, upstream of H3K9me deposition. This same interdependence has been observed for other components of the heterochromatin deposition machinery. This is typical of phenomena implicating large macromolecular structures (like heterochromatin), and pathways that take place in loops. The RNAi-heterochromatin pathway represents one such loop.

The mystery of H3K9me-siRNA interdependence was solved with the purification of the *Ago1* effector complex, known as the RNA Interference Transcriptional Silencing (RITS) complex (Verdel et al. 2004). This complex is composed of one *Ago1* bound to its siRNA, and two other proteins: the chromodomain protein *Chp1* and an adaptor protein that bridges the two, called *Tas3*. *Chp1*, by virtue of its chromodomain binding to H3K9me (Verdel et al. 2004; Schalch et al. 2009), recruits RITS to heterochromatin regions. siRNA generation is also necessary to recruit RITS to pericentric repeats (Sugiyama et al. 2005). Without the dual targeting signals of H3K9me and an siRNA, RITS cannot exert its action on heterochromatin.

What is the function of RITS? Equipped with a siRNA with homology to the reverse strand of the precursor that yielded it, and brought to the locus where these are transcribed, RITS can target the nascent precursors. As a slicer-capable member of the Argonaute family, it can then slice it. Whether slicing is necessary for subsequent processing of the precursor is still not clear; mutations in the DDH motif that constitutes the catalytic triad of the RNAse activity eliminate siRNA (Irvine et al. 2006), but this could be the result of inefficient removal of the complementary strand of the siRNA, generated by *Dcr1* processing of dsRNA. Forcing *Ago1* binding to a reporter nascent RNA, by way of a chimera with a sequence-specific RNA binding domain, bypasses the initial requirement for siRNA and results in heterochromatin deposition and siRNA generation at the targeted locus (Bühler et al. 2006). Therefore, RITS binding to the nascent RNA directs the subsequent processing steps of RNAi.

Once targeted by RITS, the nascent transcript recruits the *Rdp1* complex (RDRC) (Motamedi et al. 2004; Irvine et al. 2006), composed of *Rdp1*, an RNA helicase (*Hrr1*) and a non-canonical polyA polymerase (*Cid12*) (Motamedi et al. 2004). RDRC and RITS physically interact, indicating a direct recruitment (Motamedi et al. 2004). RDRC synthesizes RNA complementary to the targeted nascent precursor without the need for a primer (Motamedi et al. 2004), and this activity is stimulated by physical interaction with *Dcr1* (Colmenares et al. 2007). The interaction of a functional *Dcr1* with RDRC is necessary for siRNA

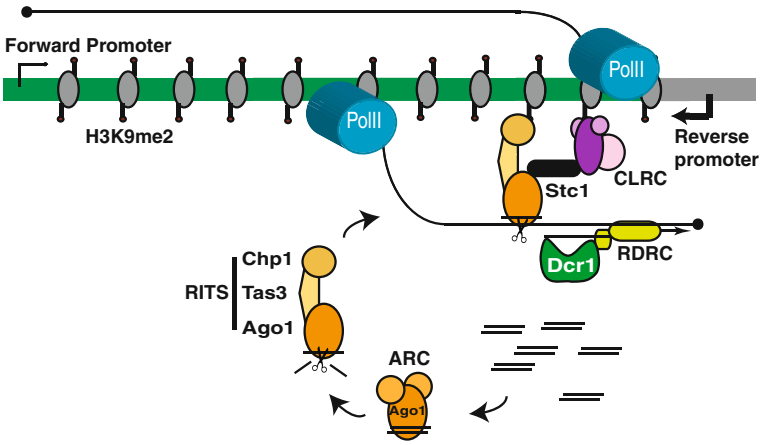


Fig. 13.4 The siRNA self-reinforcing amplification loop. See text for details

generation. RITS targeting, dsRNA generation, and siRNA processing are therefore closely coupled, acting on the nascent transcript.

Dcr1 processes the dsRNA into small dsRNA, between 20 and 26 nucleotides, with a predominant length around 24nt, both in vivo and in vitro (Motamedi et al. 2004; Colmenares et al. 2007). The small duplex RNA is captured by *Ago1* in the context of a chaperoned complex distinct from RITS, the ARC (Buker et al. 2007). Besides *Ago1*, the ARC is constituted by Arb1 and Arb2, which are conserved throughout fungi. Arb1 inhibits the slicer activity of *Ago1*. Once the ds-siRNA loaded *Ago1* is transferred to RITS, it cleaves one of the strands and strips it off the now mature siRNA. Together with *Chp1* and *Tas3*, it can now bind to the heterochromatin and target the nascent RNA, continuing the cycle. RITS can spread *in cis* via via multimerization of *Tas3*. Finally, a bridging factor, *Stc1*, mediates the recruitment of CLRC by simultaneously interacting with RITS and *Clr4* (Bayne et al. 2010) and the subsequent H3K9 methylation.

Successive rounds of targeting, dsRNA generation, and processing into siRNA leads to accumulation of siRNA in a self-reinforcing loop (Sugiyama et al. 2005) that starts with RITS binding to nascent RNA, and results in the recruitment of CLRC to pericentric repeats via *Stc1* (Fig. 13.4). A chicken-and-egg problem immediately emerges: if both siRNA and H3K9me are necessary for this initial step, which of the two comes first? Δ*clr4* mutant cells quickly and efficiently reestablish functional heterochromatin upon reintroduction of *clr4* (Partridge et al. 2007). If *Ago1* binding to its RITS partners *Tas3* and *Chp1* is disrupted, heterochromatin is not reestablished in this system. In contrast, reintroduction of *Dcr1* into a Δ*dcr1* mutant, which retains low levels of H3K9me due to the action of SHREC, leads to immediate reestablishment of heterochromatin. Therefore, without H3K9me targeting of *Ago1* the loop cannot be started, putting H3K9me at the top of the pathway. However, mutations in H3K9 that render it impossible to methylate do not affect dg siRNA levels (Gerace et al. 2010), indicating that *Clr4*

may promote siRNA generation through several mechanisms, depending on the precursor, and perhaps methylating other substrates besides H3K9. In this respect it may mediate the concerted actions of RITS, RDRC and *Dcr1* on the precursor RNA by binding to it in the context of the CLRC complex, via the CPSF-like subunit *Rik1*.

Regardless of the need for H3K9me, siRNA amplification from precursors requires some initial preexisting siRNA. Several possibilities have been proposed for the generation of this initial signal. Since the precursors are transcribed in both forward and reverse orientation, dsRNA could potentially be generated by simple hybridization of the two strands, providing *Dcr1* with a substrate without the participation of RITS and RDRC. Pericentric repeats exhibit partial inverted repeat sequences that fold into hairpins, which can be processed into siRNA by *Dcr1* (Djupedal et al. 2009). Finally, precursors are degraded in a *Dcr1* independent fashion by the action of the exosome. Some of these degraded RNAs can bind to *Ago1*, and if it can then find a reverse transcript to target, the siRNA generation cycle can commence (Halic and Moazed 2010). These primary siRNAs, or priRNAs, are derived from single-stranded transcripts by the concerted action of *Ago1* and a novel 3′–5′ exonuclease, Triman (Marasovic et al. 2013). In RDRC or *dcr1* mutants degradation products from rDNA and transcribed genes are bound to *Ago1*, but the bidirectional nature of the siRNA precursors, and the targeting contribution of H3K9me, restricts the amplification loop away from euchromatic genes and onto pericentric repeats.

Interaction with Other RNA Degradation Pathways: The Exosome Connection

The exosome is a complex of conserved RNAses and accessory proteins that degrades multiple RNA substrates (reviewed in (Chlebowski et al. 2013)) (see also Chap. 7). Its activity depends on the concerted action of several targeting mechanisms to localize and degrade aberrant RNA, byproducts of maturation of RNA, and mRNA marked for disposal. As a major component of cellular RNA surveillance mechanisms, it is not surprising that the exosome has complex and important interactions with the RNAi pathway. It would appear that the exosome both collaborates and competes with RNAi to silence their targets.

The exosome degrades the products of siRNA-mediated slicing in Drosophila (Orban and Izaurralde 2005). Consistently, in fission yeast sliced siRNA precursors can only be detected in mutants of the exosomal RNAse *rrp6*, indicating an analogous role (Irvine et al. 2006). It would seem like the exosome competes with RNAi for processing of the siRNA precursors. However, mutation of the TRAMP complex, which marks RNA through polyadenylation for degradation by the exosome, lowers siRNA levels (Bühler et al. 2007), indicating some level of collaboration. Sequencing of the siRNA bound to *Ago1* in TRAMP mutants revealed the accumulation of rRNA- and tRNA-derived siRNA, suggesting that

exosome degradation of non-pericentric transcripts precluded their entry into the RNAi cycle (Buhler et al. 2008). Perhaps the role of TRAMP is to prevent these spurious RNA from overwhelming the RNAi machinery.

At the same time, the exosome and exosome-related RNA surveillance mechanisms have a role in heterochromatin deposition in their own right. Meiotic mRNAs are transcriptionally and post-transcriptionally silenced in vegetative cells, with many meiotic genes showing H3K9me. RITS localizes to these loci, but H3K9me deposition appears independent of RNAi. Their post-transcriptional silencing is often the result of binding of the Mmi1/Red1 complex to a specific sequence present in meiotic mRNA. Strikingly, introducing this sequence in reporter genes induces H3K9me deposition, and deletion of *mmi1* or *red1*, as well as *rrp6*, led to loss of H3K9me at meiotic genes (Zofall et al. 2012). Red1 interacts physically with *Clr4*, providing a mechanistic link.

Another RNA surveillance factor linked to the exosome, *Mlo3*, interacts with both RITS and *Clr4* (Zhang et al. 2011). *Mlo3* is methylated by *Clr4*, and this methylation is necessary for siRNA generation in H3K9R mutants. Perhaps *Mlo3* is the non-histone substrate by which *Clr4* can induce siRNA generation independently of H3K9me (Gerace et al. 2010). While *Mlo3* acts with TRAMP to target RNA to the exosome, it could act in concert with *Clr4* to channel the precursors to RNAi instead (Zhang et al. 2011). Mutation of *Mlo3* or TRAMP factors rescues heterochromatin in RNAi mutants (Reyes-Turcu et al. 2011). Therefore, the ncRNA that accumulate in these mutants can recruit heterochromatin independently of RNAi. A recent study of the interaction between heterochromatin and diverse RNA surveillance factors revealed the existence of a complex network of RNA binding factors, polyA polymerases, spliceosomal factors and exosome components centered in Red1 and the helicase Mtl1 that directs the processing of multiple targets by the exosome and RNAi (Lee et al. 2013) and the subsequent heterochromatin deposition.

Swi6 also collaborates with the exosome to process heterochromatic transcripts. HP1 family members can bind to RNA through their hinge domain (between the chromo and chromoshadow domains) (Maison et al. 2002; Keller et al. 2012). ncRNA binding releases *Swi6* from H3K9me, and is subsequently transferred to the TRAMP complex for degradation by the exosome. This process effectively keeps heterochromatin-derived transcripts away from the translation machinery(Keller et al. 2012).

Inheritance of Heterochromatin Through Cell Division

Cell Cycle Regulation of Heterochromatin

Transcription of the siRNA precursor triggers a frenzy of processing activities that result in heterochromatin deposition and transcriptional silencing. This led to a paradox that contradicted decades of research in heterochromatin: the silenced

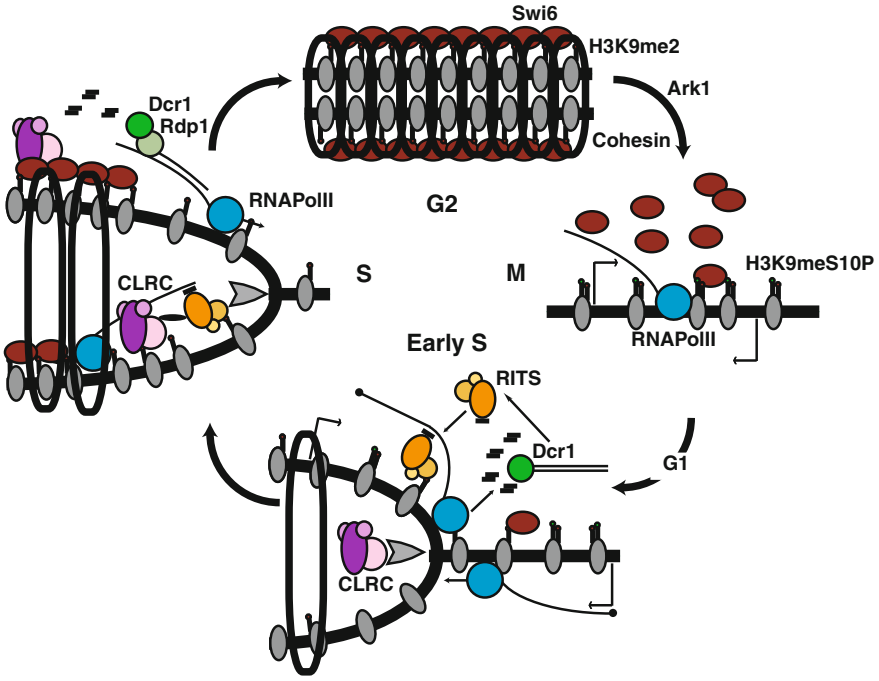


Fig. 13.5 Cell cycle regulation of RNAi and heterochromatin inheritance. See text for details

regions were not, after all, completely silent. HP1 binding is highly dynamic, perhaps allowing for some precursor transcription (Fischle et al. 2005), and a portion of the observed silencing could be post-transcriptional, or even co-transcriptional (Bühler et al. 2006). But transcription run-on and RNA pol II ChIP experiments are unequivocal: wild type heterochromatin is transcriptionally silent. Where does the transcription necessary for siRNA generation come from?

The solution to this paradox is that it's not the "where" but the "when" that matters. Most experiments are carried out in cells growing in logarithmic phase in liquid culture. These are composed of unsynchronized populations of cells in all stages of the cell cycle. The simultaneous presence of siRNA and silent heterochromatin could be due to each of these being restricted to a particular step in the cell cycle.

S. pombe spends about 70 % of the time in G2. Mitosis is immediately followed by S-phase, so G1 is extremely brief. When cultures were synchronized by hydroxyurea treatment (synchronizing at the start of S-phase) or by a temperature sensitive mutation in *cdc25* (synchronizing at G2) and then released, heterochromatin and RNAi were observed to follow a cycle closely coupled to the cell cycle (Chen et al. 2008; Kloc et al. 2008) (Fig. 13.5). G2 heterochromatin is fully silent and shows high levels of H3K9me and *Swi6*. Critically, H3S10 phosphorylation by Aurora kinase at the onset of mitosis disrupts *Swi6* binding (Fischle et al. 2005),

allowing for the transcription of the siRNA precursors as detected by RNA levels and RNA pol II ChIP. After some initial accumulation of precursor, siRNA start to accumulate in late mitosis, reaching maximum levels coinciding with early S-phase. H3K9me levels fall slightly during S-phase but then recover, indicating that incorporation of naïve nucleosomes during DNA replication dilutes H3K9me transiently. Precursor levels, and then siRNA levels fall after the reestablishment of heterochromatin, and by G2 are completely absent. ChIP of CLRC and RITS complex components show a similar pattern.

Thus, a close coupling of transcription and RNAi processing of the siRNA precursors with the cell cycle prepares heterochromatin for reestablishment after incorporation of naïve nucleosomes. In this model, the initial siRNA that trigger the self-reinforcing loop can be newly generated from priRNA or left over from a previous S-phase, and the remaining H3K9me in heterochromatin would target RITS to direct *Clr4* activity to the naïve nucleosomes, resulting in the maintenance of heterochromatin through mitosis. Heterochromatin is therefore not a static structure, but palpitates once per cell cycle to provide the sequence specificity that allows its inheritance.

Regulation of Heterochromatic DNA Replication

This close coupling of RNAi processing with S-phase suggests that DNA replication dilutes heterochromatin, but is also strongly implicated in its inheritance. The pericentric repeats in fission yeast are replicated early, in contrast with heterochromatin in other organisms where heterochromatic DNA is late replicating. This is due to the activity of *Swi6* that stimulates loading of the pre-replication complex subunit Sld3 to origins contained in the pericentric repeats. DNA replication origins have been mapped by plasmid-maintenance assays and genome-wide approaches (Smith et al. 1995; Segurado et al. 2003). When compared with the location of the siRNA precursors, these origins are located in the untranscribed regions between siRNA clusters (Zaratiegui et al. 2011), never within them (Fig. 13.1). Such a striking arrangement points to an unexpected coordination between RNAi and DNA replication.

How this coordination is carried out, and how it facilitates heterochromatin deposition, is still not clear. Components of the CLRC complex interact physically with *Cdc20*, the catalytic subunit of the DNA pol ε, placing it in direct contact with the advancing replisome (Li et al. 2011). This raises the interesting possibility that *Clr4* is reestablishing heterochromatin in the wake of the replication fork by methylating H3K9 in the naïve nucleosomes as they are incorporated in the daughter chromatids.

Analysis of DNA replication intermediates of pericentric repeats reveals that the progression of the replication fork through these regions is problematic, with visible fork arrest signals (Zaratiegui et al. 2011). This could be due to collisions between advancing DNA and RNA polymerases. RNAi processing of the precursors alleviates the situation by releasing RNA pol II, allowing resumption of

DNA replication. In *dcr1* mutants the arrested forks stall, losing the replisome, and engage the homologous recombination (HR) machinery to restart replication. Concomitantly, CLRC interaction with *Cdc20* is partially lost. RNAi could therefore mediate replication-coupled heterochromatin reestablishment by loading CLRC via *Stc1* onto the replisome. Alternatively, CLRC might be loaded by *Swi6* at heterochromatic replication origins, and RNAi would simply prevent the loss of the CLRC-equipped replisome by solving the collisions with RNA pol II.

Interaction with Homologous Recombination

Replication of pericentric repeats, even in the wild type, is troublesome. A survey of the DNA damage chromatin signal gamma-H2A in undisturbed cultures revealed that this mark accumulated over the pericentric repeats, with a profile that mirrored siRNA density (Rozenzhak et al. 2010). This is probably due to the presence of replication impediments. Active transcription of siRNA precursors surrounding the centromeric origins could pose such an impediment. In the absence of RNAi, resolution of replication fork arrest at the repeats requires mitotic HR (Zaratiegui et al. 2011), which restarts replication by using homologous sequences. HR and RNAi can therefore redundantly restart replication of the arrested forks. Consistently, RNAi mutants exhibit synthetic sickness with factors involved in replication fork protection (Roguev et al. 2008; Zaratiegui et al. 2011), indicating that the lesions that arise in RNAi mutants require the action of HR to prevent catastrophic damage. The regulation between RNAi and HR may happen both ways: a recent high-throughput screen in *Neurospora* looking for factors necessary for generation of qiRNA (a type of siRNA induced by DNA damage) yielded HR genes (Zhang et al. 2013) (see Chap. 11). HR is therefore also capable of influencing RNAi.

HR at repetitive DNA is a dangerous game. The homologous sequence that serves as a surrogate template for the restart of replication is usually restricted to the sister chromatid, but if cohesion is compromised (as it often is in heterochromatin mutants) the search for homology could involve non-allelic copies of the repetitive sequences. In the best case scenario this could lead to gene conversion between different repeats, or changes in their copy number. This is the current model that explains the growth of centromeric satellites and their sequence homogeneity. But non-allelic recombination can also cause gross chromosomal rearrangements, and if the recombination intermediates are not properly resolved they can lead to fragile anaphase bridges that prevent proper chromosome segregation and cause further DNA damage (Sofueva et al. 2011).

By preempting the role of HR in resolving replication conflicts in pericentric repeats, RNAi acts as a protector of genome integrity. In this view, the long known function of heterochromatin in the maintenance of genome integrity is explained by its role in facilitating seamless replication of repetitive DNA. A corollary of this

model is that the cause of genome instability in mutants that lose heterochromatin is caused by the replication fork impediments that fall back on HR for resolution.

The pericentric sequences are arranged as an inverted repeat around the central core, and fork arrest at inverted repeats leads to mitotic HR that can cause rearrangements and even deletion of the intervening sequences (Lambert et al. 2005, 2010). The pericentric regions of both arms of each centromere cluster together in interphase (Kniola et al. 2001), probably facilitating the non-allelic HR. Non-allelic recombination and gene conversion between pericentric repeats may be the cause of the size variability in centromeres from different isolates, their structural homogeneity within each centromere, and the occasional inversion of the central core (Steiner et al. 1993; Nakamura et al. 2008).

RNAi is not the only heterochromatin machinery with a role in managing DNA replication. Mutants in other heterochromatin factors, like *clr4* and *swi6*, also show synthetic sickness with fork protectors, and increases in rearrangements at the centromere and chromosome loss (Li et al. 2013), perhaps indicating additional sources of fork instability within the repeats. Heterochromatin in general therefore has a role as a safeguard for correct DNA replication of repetitive DNA, explaining its function as a protector of genome integrity.

NcRNA at the Boundary Elements

H3K9me deposition is restricted to the pericentric repeats, and excluded from the central core and the flanking euchromatin. Present at almost every boundary region are tRNA genes, individually or organized in clusters (Kuhn et al. 1991). The exception is the boundary between the right pericentric region and euchromatin at chromosome 1 (IRC1R), where an inverted repeat element, also present in the left boundary (IRC1L) and in centromere 3 (IRC3L/R), appears to act as the barrier (Cam et al. 2005).

tRNA Gene Clusters

tRNA genes display insulator functions in many organisms (Van Bortle and Corces 2012) (see also Chap. 10). This function requires binding of the transcription factor TFIIIC and RNA pol III, and this is recapitulated in the $tRNA^{ala}$ that separates the otr from euchromatin in the left arm of centromere 1 (Scott et al. 2006). The exact mechanism that interrupts *in cis* spreading of heterochromatin is not known, but may involve specific counteracting chromatin modifications that localize to tRNA genes. Alternatively, the chromatin structure at tRNA genes may not be conducive to spreading: in *S. cerevisiae* tRNA genes have a prominent nucleosome-free region (Kumar and Bhargava 2013) that could deprive the

spreading heterochromatin of a docking platform to continue the cycle of modification of adjacent nucleosomes.

TFIIIC binding regions have a general function in the organization of subnuclear architecture (Noma et al. 2006). Genome-wide approaches showed that TFIIIC binding to scattered elements, among them tRNA genes, mediated their clustering in distinct subnuclear bodies localized to the nuclear periphery. One such sequence, an inverted repeat, also serves as the boundary elements that prevent spreading of mating type heterochromatin. The two functions could be functionally related, and tethering to the nuclear periphery may create separate chromatin domains on either side of the TFIIIC binding element.

IRC1R and the Borderline ncRNA

IRC1 and IRC3 do not show TFIIIC binding, so their barrier activity must involve other mechanisms (Noma et al. 2006). Instead, these regions appear enriched for euchromatic marks and are actively transcribed by RNA pol II. The ncRNA is processed into siRNA, but strikingly this does not induce the deposition of heterochromatin (Cam et al. 2005; Noma et al. 2006). Recently a ncRNA, termed *borderline*, originating around IRC1R was shown to be responsible for the barrier activity of this region (Keller et al. 2013). Much like the dg and dh ncRNA, the transcript size is heterogeneous, and *Dcr1* processes them into a novel class of small RNA called brdrRNA. However these siRNA-like molecules are not incorporated into *Ago1*, explaining the absence of heterochromatin in this element. How *borderline* can be efficiently processed by *Dcr1* into brdrRNA without the participation of *Ago1* in a self-reinforcing loop is an open question that will probably yield new insights into the mechanisms of RNAi.

The barrier activity of *borderline* requires the RNA binding function of *Swi6*. Since RNA binding to *Swi6* removes it from H3K9me, perhaps *borderline* or brdrRNA binding results in active disengagement of the spreading *Swi6* multimer as it reaches IRC1R. Interestingly, the ncRNA is not necessary, and it can be substituted with an actively transcribed protein coding gene that is also processed into brdrRNA. The IRC1L and IRC3L/R repeats, being present together with tRNA gene clusters, probably collaborate with them setting the heterochromatin boundaries. But a *Swi6* mutant that loses RNA binding only shows weak ectopic spreading of heterochromatin beyond IRC1L, and none beyond IRC3L/R, indicating that as long as there's at least one tRNA gene some barrier activity is retained.

Concluding Remarks

Contrary to long-held beliefs, heterochromatin is neither silent nor static, but a hotbed of transcriptional activity with highly dynamic behavior. The detailed picture of ncRNA transcription and processing that arises from the study of *S. pombe* pericentric repeats is surely but one example of the control of heterochromatin by ncRNA. Other model systems, like *Neurospora*, *S. japonicus* and the satellite arrays in mammalian centromeres, will likely reveal the conserved principles behind this phenomenon.

References

Al-Sady B, Madhani HD, Narlikar GJ (2013) Division of labor between the chromodomains of HP1 and Suv39 methylase enables coordination of heterochromatin spread. Mol Cell 51:80–91. doi:10.1016/j.molcel.2013.06.013

Allshire R, Javerzat J, Redhead N, Cranston G (1994) Position effect variegation at fission yeast centromeres. Cell 76:157–169

Alper BJ, Job G, Yadav RK et al (2013) Sir2 is required for Clr4 to initiate centromeric heterochromatin assembly in fission yeast. EMBO J 32:2321–2335. doi:10.1038/emboj.2013.143

Ayoub N, Goldshmidt I, Cohen A (1999) Position effect variegation at the mating-type locus of fission yeast: a cis-acting element inhibits covariegated expression of genes in the silent and expressed domains. Genetics 152:495–508

Bannister AJ, Zegerman P, Partridge JF et al (2001) Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature 410:120–124. doi:10.1038/35065138

Barraud P, Emmerth S, Shimada Y et al (2011) An extended dsRBD with a novel zinc-binding motif mediates nuclear retention of fission yeast Dicer. EMBO J 30:4223–4235. doi:10.1038/emboj.2011.300

Baum M, Ngan VK, Clarke L (1994) The centromeric K-type repeat and the central core are together sufficient to establish a functional *Schizosaccharomyces pombe* centromere. Mol Biol Cell 5:747–761

Bayne EH, White SA, Kagansky A et al (2010) *Stc1*: a critical link between RNAi and chromatin modification required for heterochromatin integrity. Cell 140:666–677. doi:10.1016/j.cell.2010.01.038

Bernard P, Maure J, Partridge J et al (2001) Requirement of heterochromatin for cohesion at centromeres. Science 294:2539–2542

Buhler M, Spies N, Bartel DP, Moazed D (2008) TRAMP-mediated RNA surveillance prevents spurious entry of RNAs into the *Schizosaccharomyces pombe* siRNA pathway. Nat Struct Mol Biol 15:1015–1023. doi:10.1038/nsmb.1481

Buker SM, Iida T, Buhler M et al (2007) Two different argonaute complexes are required for siRNA generation and heterochromatin assembly in fission yeast. Nat Struct Mol Biol 14:200–207. doi:10.1038/nsmb1211

Buscaino A, Lejeune E, Audergon P et al (2013) Distinct roles for Sir2 and RNAi in centromeric heterochromatin nucleation, spreading and maintenance. EMBO J. doi:10.1038/emboj.2013.72

Bühler M, Haas W, Gygi S, Moazed D (2007) RNAi-dependent and-independent RNA turnover mechanisms contribute to heterochromatic gene silencing. Cell 129:707

Bühler M, Verdel A, Moazed D (2006) Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. Cell 125:873–886. doi:10.1016/j.cell.2006.04.025

Cam H, Sugiyama T, Chen E et al (2005) Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. Nat Genet 37:809–819

Cam HP, Noma K-I, Ebina H et al (2008) Host genome surveillance for retrotransposons by transposon-derived proteins. Nature 451:431–436. doi:10.1038/nature06499

Cambareri EB, Aisner R, Carbon J (1998) Structure of the chromosome VII centromere region in *neurospora* crassa: degenerate transposons and simple repeats. Mol Cell Biol 18(9):5465–5477

Canzio D, Chang EY, Shankar S et al (2011) Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly. Mol Cell 41:67–81. doi:10.1016/j.molcel.2010.12.016

Carmichael JB, Stoica C, Parker H et al (2006) RNA interference effector proteins localize to mobile cytoplasmic puncta in *Schizosaccharomyces pombe*. Traffic 7:1032–1044. doi:10.1111/j.1600-0854.2006.00441.x

Chen E, Zhang K, Nicolas E et al (2008) Cell cycle control of centromeric repeat transcription and heterochromatin assembly. Nature. doi:10.1038/nature06561

Chinen M, Morita M, Fukumura K, Tani T (2010) Involvement of the Spliceosomal U4 small nuclear RNA in heterochromatic gene silencing at fission yeast centromeres. J Biol Chem 285:5630–5638. doi:10.1074/jbc.M109.074393

Chlebowski A, Lubas M, Jensen TH, Dziembowski A (2013) RNA decay machines: the exosome. Biochim Biophys Acta 1829:552–560. doi:10.1016/j.bbagrm.2013.01.006

Clarke L, Amstutz H, Fishel B, Carbon J (1986) Analysis of centromeric DNA in the fission yeast *Schizosaccharomyces pombe*. Proc Natl Acad Sci USA 83:8253–8257

Clarke L, Baum MP (1990) Functional analysis of a centromere from fission yeast: a role for centromere-specific repeated DNA sequences. Mol Cell Biol 10(5):1863–1872

Colmenares SU, Buker SM, Buhler M et al (2007) Coupling of double-stranded RNA synthesis and siRNA generation in fission yeast RNAi. Mol Cell 27:449–461. doi:10.1016/j.molcel.2007.07.007

Djupedal I (2005) RNA Pol II subunit *Rpb7* promotes centromeric transcription and RNAi-directed chromatin silencing. Genes Dev 19:2301–2306. doi:10.1101/gad.344205

Djupedal I, Kos-Braun IC, Mosher RA et al (2009) Analysis of small RNA in fission yeast; centromeric siRNAs are potentially generated through a structured RNA. EMBO J. doi:10.1038/emboj.2009.351

Drinnenberg IA, Fink GR, Bartel DP (2011) Compatibility with killer explains the rise of RNAi-deficient fungi. Science 333:1592. doi:10.1126/science.1209575

Du J, Zhong X, Bernatavichute YV et al (2012) Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell 151:167–180. doi:10.1016/j.cell.2012.07.034

Ekwall K, Cranston G, Allshire R (1999) Fission yeast mutants that alleviate transcriptional silencing in centromeric flanking repeats and disrupt chromosome segregation. Genetics 153:1153–1169

Ekwall K, Javerzat JP, Lorentz A et al (1995) The chromodomain protein *Swi6*: a key component at fission yeast centromeres. Science 269:1429–1431

Ekwall K, Ruusala T (1994) Mutations in *rik1*, *clr2*, *clr3* and *clr4* genes asymmetrically derepress the silent mating-type loci in fission yeast. Genetics 136:53–64

Emmerth S, Schober H, Gaidatzis D et al (2010) Nuclear retention of fission yeast dicer is a prerequisite for RNAi-mediated heterochromatin assembly. Dev Cell 18:102–113. doi:10.1016/j.devcel.2009.11.011

Fischer T, Cui B, Dhakshnamoorthy J et al (2009) Diverse roles of HP1 proteins in heterochromatin assembly and functions in fission yeast. Proc Natl Acad Sci USA 106:8998–9003. doi:10.1073/pnas.0813063106

Fischle W, Tseng BS, Dormann HL et al (2005) Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. Nature 438:1116–1122. doi:10.1038/nature04219

Fishel B, Amstutz H, Baum M et al (1988) Structural organization and functional analysis of centromeric DNA in the fission yeast *Schizosaccharomyces pombe*. Mol Cell Biol 8:754–763

Folco HD, Pidoux AL, Urano T, Allshire RC (2008) Heterochromatin and RNAi are required to establish CENP-A chromatin at centromeres. Science 319:94–97. doi:10.1126/science.1150944

Garcia JF, Dumesic PA, Hartley PD et al (2010) Combinatorial, site-specific requirement for heterochromatic silencing factors in the elimination of nucleosome-free regions. Genes Dev 24:1758–1771. doi:10.1101/gad.1946410

Gerace EL, Halic M, Moazed D (2010) The methyltransferase activity of Clr4Suv39 h triggers RNAi independently of histone H3K9 methylation. Mol Cell 39:360–372. doi:10.1016/j.molcel.2010.07.017

Grewal SI, Klar AJ (1997) A recombinationally repressed region between mat2 and mat3 loci shares homology to centromeric repeats and regulates directionality of mating-type switching in fission yeast. Genetics 146:1221–1238

Halic M, Moazed D (2010) Dicer-independent primal RNAs trigger RNAi and heterochromatin formation. Cell 140:504–516. doi:10.1016/j.cell.2010.01.019

Hall IM, Noma K-I, Grewal SIS (2003) RNA interference machinery regulates chromosome dynamics during mitosis and meiosis in fission yeast. Proc Natl Acad Sci USA 100:193–198. doi:10.1073/pnas.232688099

Hall IM, Shankaranarayana GD, Noma K-I et al (2002) Establishment and maintenance of a heterochromatin domain. Science 297:2232–2237. doi:10.1126/science.1076466

Honda S, Lewis ZA, Shimada K et al (2012) Heterochromatin protein 1 forms distinct complexes to direct histone deacetylation and DNA methylation. Nat Struct Mol Biol 19:471–477. doi:10.1038/nsmb.2274

Hong E-JE, Villén J, Moazed D (2005) A cullin E3 ubiquitin ligase complex associates with *Rik1* and the *Clr4* histone H3-K9 methyltransferase and is required for RNAi-mediated heterochromatin formation. RNA Biol 2:106–111. doi:10.4161/rna.2.3.2131

Horn P, Bastie J, Peterson C (2005) A *Rik1*-associated, cullin-dependent E3 ubiquitin ligase is essential for heterochromatin formation. Genes Dev 19:1705–1714

Irvine DV, Zaratiegui M, Tolia NH et al (2006) Argonaute slicing is required for heterochromatic silencing and spreading. Science 313:1134–1137. doi:10.1126/science.1128813

Javerzat JP, McGurk G, Cranston G et al (1999) Defects in components of the proteasome enhance transcriptional silencing at fission yeast centromeres and impair chromosome segregation. Mol Cell Biol 19:5155–5165

Jia S, Kobayashi R, Grewal S (2005) Ubiquitin ligase component *Cul4* associates with *Clr4* histone methyltransferase to assemble heterochromatin. Nat Cell Biol 7:1007–1013

Jia S, Noma K-I, Grewal SIS (2004a) RNAi-independent heterochromatin nucleation by the stress-activated ATF/CREB family proteins. Science 304:1971–1976. doi:10.1126/science.1099035

Jia S, Yamada T, Grewal S (2004b) Heterochromatin regulates cell type-specific long-range chromatin interactions essential for directed recombination. Cell 119:469–480

Kagansky A, Folco HD, Almeida R et al (2009) Synthetic heterochromatin bypasses RNAi and centromeric repeats to establish functional centromeres. Science 324:1716–1719. doi:10.1126/science.1172026

Kato H, Goto D, Martienssen R et al (2005) RNA polymerase II is required for RNAi-dependent heterochromatin assembly. Science 309:467–469

Keller C, Adaixo R, Stunnenberg R et al (2012) HP1*Swi6* mediates the recognition and destruction of heterochromatic RNA transcripts. Mol Cell 47:215–227. doi:10.1016/j.molcel.2012.05.009

Keller C, Kulasegaran-Shylini R, Shimada Y et al (2013) Noncoding RNAs prevent spreading of a repressive histone mark. Nat Struct Mol Biol. doi:10.1038/nsmb.2619

Kipling D, Warburton PE (1997) Centromeres, CENP-B and tigger too. Trends Genet 13:141–145

Klar AJ, Bonaduce MJ (1991) *Swi6*, a gene required for mating-type switching, prohibits meiotic recombination in the mat2-mat3 “cold spot” of fission yeast. Genetics 129:1033–1042

Kloc A, Zaratiegui M, Nora E, Martienssen R (2008) RNA interference guides histone modification during the S phase of chromosomal replication. Curr Biol 18:490–495. doi:10.1016/j.cub.2008.03.016

Kniola B, O'Toole E, McIntosh JR et al (2001) The domain structure of centromeres is conserved from fission yeast to humans. Mol Biol Cell 12:2767–2775. doi:10.1091/mbc.12.9.2767

Kuhn RM, Clarke L, Carbon J (1991) Clustered tRNA genes in *Schizosaccharomyces pombe* centromeric DNA sequence repeats. Proc Natl Acad Sci USA 88:1306–1310

Kumar Y, Bhargava P (2013) A unique nucleosome arrangement, maintained actively by chromatin remodelers facilitates transcription of yeast tRNA genes. BMC Genom 14:402. doi:10.1186/1471-2164-14-402

Lambert S, Mizuno K, Blaisonneau J et al (2010) Homologous recombination restarts blocked replication forks at the expense of genome rearrangements by template exchange. Mol Cell 39:346–359. doi:10.1016/j.molcel.2010.07.015

Lambert S, Watson A, Sheedy DM et al (2005) Gross chromosomal rearrangements and elevated recombination at an inducible site-specific replication fork barrier. Cell 121:689–702. doi:10.1016/j.cell.2005.03.022

Lee H-C, Li L, Gu W et al (2010) Diverse pathways generate microRNA-like RNAs and Dicer-independent small interfering RNAs in fungi. Mol Cell 38:803–814. doi:10.1016/j.molcel.2010.04.005

Lee NN, Chalamcharla VR, Reyes-Turcu F et al (2013) Mtr4-like protein coordinates nuclear RNA processing for heterochromatin assembly and for telomere maintenance. Cell. doi:10.1016/j.cell.2013.10.027

Lehnertz B, Ueda Y, Derijck AAHA et al (2003) Suv39 h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr Biol 13:1192–1200. doi:10.1016/S0960-9822(03)00432-9

Li F, Goto DB, Zaratiegui M et al (2005) Two novel proteins, *dos1* and *dos2*, interact with *rik1* to regulate heterochromatic RNA interference and histone modification. Curr Biol 15:1448–1457. doi:10.1016/j.cub.2005.07.021

Li F, Martienssen R, Cande WZ (2011) Coordination of DNA replication and histone modification by the *Rik1-Dos2* complex. Nature 475:244–248. doi:10.1038/nature10161

Li P-C, Petreaca RC, Jensen A et al (2013) Replication fork stability is essential for the maintenance of centromere integrity in the absence of heterochromatin. Cell Reports. doi:10.1016/j.celrep.2013.02.007

Lorentz A, Ostermann K, Fleck O, Schmidt H (1994) Switching gene *swi6*, involved in repression of silent mating-type loci in fission yeast, encodes a homologue of chromatin-associated proteins from Drosophila and mammals. Gene 143:139–143. doi:10.1016/0378-1119(94)90619-X

Maison C, Bailly D, Peters A et al (2002) Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. Nat Genet 30:329–334

Marasovic M, Zocco M, Halic M (2013) Argonaute and triman generate dicer-independent priRNAs and mature siRNAs to initiate heterochromatin formation. Mol Cell 52:173–183. doi:10.1016/j.molcel.2013.08.046

Marina DB, Shankar S, Natarajan P et al (2013) A conserved ncRNA-binding protein recruits silencing factors to heterochromatin through an RNAi-independent mechanism. Genes Dev 27:1851–1856. doi:10.1101/gad.226019.113

Marschall LG, Clarke L (1995) A novel cis-acting centromeric DNA element affects *S. pombe* centromeric chromatin structure at a distance. J Cell Biol 128:445–454

Melters DP, Bradnam KR, Young HA et al (2013) Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. Genome Biol 14:R10. doi:10.1007/s00412-010-0278-5

Motamedi M, Verdel A, Colmenares S et al (2004) Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. Cell 119:789–802

Motamedi MR, Hong E-JE, Li X et al (2008) HP1 proteins form distinct complexes and mediate heterochromatic gene silencing by nonoverlapping mechanisms. Mol Cell 32:778–790. doi:10.1016/j.molcel.2008.10.026

Nakamura K-I, Okamoto A, Katou Y et al (2008) Rad51 suppresses gross chromosomal rearrangement at centromere in *Schizosaccharomyces pombe*. EMBO J 27:3036–3046. doi:10.1038/emboj.2008.215

Nakaseko Y, Adachi Y, Funahashi S et al (1986) Chromosome walking shows a highly homologous repetitive sequence present in all the centromere regions of fission yeast. EMBO J 5:1011–1021

Nakaseko Y, Kinoshita N, Yanagida M (1987) A novel sequence common to the centromere regions of *Schizosaccharomyces pombe* chromosomes. Nucleic Acids Res 15:4705–4715

Nakayama J, Rice JC, Strahl BD et al (2001) Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science 292:110–113. doi:10.1126/science.1060118

Nielsen ML, Hermansen TD, Aleksenko A (2001) A family of DNA repeats in *aspergillus nidulans* has assimilated degenerated retrotransposons. Mol Genet Genomics 265:883–887

Noma K, Cam H, Maraia R, Grewal S (2006) A role for TFIIIC transcription factor complex in genome organization. Cell 125:859–872

Orban T, Izaurralde E (2005) Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. RNA 11:459–469. doi:10.1261/rna.7231505

Partridge JF, DeBeauchamp JL, Kosinski AM et al (2007) Functional separation of the requirements for establishment and maintenance of centromeric heterochromatin. Mol Cell 26:593–602. doi:10.1016/j.molcel.2007.05.004

Partridge JF, Scott KSC, Bannister AJ et al (2002) cis-acting DNA from fission yeast centromeres mediates histone H3 methylation and recruitment of silencing factors and cohesin to an ectopic site. Curr Biol 12:1652–1660

Peters AH, O'Carroll D, Scherthan H et al (2001) Loss of the Suv39 h histone methyltransferases impairs mammalian heterochromatin and genome stability. Cell 107:323–337

Polizzi C, Clarke L (1991) The chromatin structure of centromeres from fission yeast: differentiation of the central core that correlates with function. J Cell Biol 112:191–201

Quénet D, Dalal Y (2012) The CENP-A nucleosome: a dynamic structure and role at the centromere. Chromosome Res 20:465–479. doi:10.1007/s10577-012-9301-4

Rea S, Eisenhaber F, O'Carroll D et al (2000) Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature 406:593–599. doi:10.1038/35020506

Reinhart BJ, Bartel DP (2002) Small RNAs correspond to centromere heterochromatic repeats. Science 297:1831. doi:10.1126/science.1077183

Reyes-Turcu FE, Zhang K, Zofall M et al (2011) Defects in RNA quality control factors reveal RNAi-independent nucleation of heterochromatin. Nat Struct Mol Biol 18:1132–1138. doi:10.1038/nsmb.2122

Rhind N, Chen Z, Yassour M et al (2011) Comparative functional genomics of the fission yeasts. Science 332:930–936. doi:10.1126/science.1203357

Roguev A, Bandyopadhyay S, Zofall M et al (2008) Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. Science 322:405–410. doi:10.1126/science.1162609

Rozenzhak S, Mejía-Ramírez E, Williams JS et al (2010) Rad3 decorates critical chromosomal domains with gammaH2A to protect genome integrity during S-Phase in fission yeast. PLoS Genet 6:e1001032. doi:10.1371/journal.pgen.1001032

Sadaie M, Iida T, Urano T, Nakayama J (2004) A chromodomain protein, *Chp1*, is required for the establishment of heterochromatin in fission yeast. EMBO J 23:3825–3835

Schalch T, Job G, Noffsinger VJ et al (2009) High-affinity binding of *Chp1* chromodomain to K9 methylated histone H3 is required to establish centromeric heterochromatin. Mol Cell 34:36–46. doi:10.1016/j.molcel.2009.02.024

Scott K, Merrett S, Willard H (2006) A heterochromatin barrier partitions the fission yeast centromere into discrete chromatin domains. Curr Biol 16:119–129

Segurado M, de Luis A, Antequera F (2003) Genome-wide distribution of DNA replication origins at A + T-rich islands in *Schizosaccharomyces pombe*. EMBO Rep 4:1048–1053. doi:10.1038/sj.embor.embor7400008

Sigova A, Rhind N, Zamore PD (2004) A single argonaute protein mediates both transcriptional and posttranscriptional silencing in *Schizosaccharomyces pombe*. Genes Dev 18:2359–2367. doi:10.1101/gad.1218004

Simmer F, Buscaino A, Kos-Braun IC et al (2010) Hairpin RNA induces secondary small interfering RNA synthesis and silencing in trans in fission yeast. EMBO Rep 11:112–118. doi:10.1038/embor.2009.273

Smith J, Caddle M, Bulboaca G et al (1995) Replication of centromere II of *Schizosaccharomyces pombe*. Mol Cell Biol 15:5165

Smith KM, Galazka JM, Phatale PA et al (2012) Centromeres of filamentous fungi. Chromosome Res 20:635–656. doi:10.1007/s10577-012-9290-3

Sofueva S, Osman F, Lorenz A et al (2011) Ultrafine anaphase bridges, broken DNA and illegitimate recombination induced by a replication fork barrier. Nucleic Acids Res 39:6568–6584. doi:10.1093/nar/gkr340

Steiner NC, Hahnenberger KM, Clarke L (1993) Centromeres of the fission yeast *Schizosaccharomyces pombe* are highly variable genetic loci. Mol Cell Biol 13:4578–4587

Sugiyama T, Cam H, Verdel A et al (2005) RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production. Proc Natl Acad Sci U S A 102:152–157

Sugiyama T, Cam HP, Sugiyama R et al (2007) SHREC, an effector complex for heterochromatic transcriptional silencing. Cell 128:491–504. doi:10.1016/j.cell.2006.12.035

Takahashi K, Chen E, Yanagida M (2000) Requirement of Mis6 centromere connector for localizing a CENP-A-like protein in fission yeast. Science 288:2215–2219

Thon G, Cohen A, Klar AJ (1994) Three additional linkage groups that repress transcription and meiotic recombination in the mating-type region of *Schizosaccharomyces pombe*. Genetics 138:29–38

Thon G, Hansen K, Altes S et al (2005) The Clr7 and Clr8 directionality factors and the Pcu4 cullin mediate heterochromatin formation in the fission yeast *Schizosaccharomyces pombe*. Genetics 171:1583–1595

Van Bortle K, Corces VG (2012) tDNA insulators and the emerging role of TFIIIC in genome organization. Transcription 3:277–284. doi:10.4161/trns.21579

Verdel A, Jia S, Gerber S et al (2004) RNAi-mediated targeting of heterochromatin by the RITS complex. Science 303:672–676

Volpe T, Kidner C, Hall I et al (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science 297:1833–1837

Volpe T, Schramke V, Hamilton G et al (2003) RNA interference is required for normal centromere function in fission yeast. Chromosome Res 11:137–146

Wood V, Gwilliam R, Rajandream M-A et al (2002) The genome sequence of *Schizosaccharomyces pombe*. Nature 415:871–880. doi:10.1038/nature724

Woolcock KJ, Gaidatzis D, Punga T, Buhler M (2011) Dicer associates with chromatin to repress genome activity in *Schizosaccharomyces pombe*. Nat Struct Mol Biol 18:94–99. doi:10.1038/nsmb.1935

Woolcock KJ, Stunnenberg R, Gaidatzis D et al (2012) RNAi keeps *Atf1*-bound stress response genes in check at nuclear pores. Genes Dev 26:683–692. doi:10.1101/gad.186866.112

Xhemalce B, Kouzarides T (2010) A chromodomain switch mediated by histone H3 Lys 4 acetylation regulates heterochromatin assembly. Genes Dev 24:647–652. doi:10.1101/gad.1881710

Yamada T, Fischle W, Sugiyama T et al (2005) The nucleation and maintenance of heterochromatin by a histone deacetylase in fission yeast. Mol Cell 20:173–185. doi:10.1016/j.molcel.2005.10.002

Yu R, Jih G, Iglesias N, Moazed D (2013) Determinants of heterochromatic siRNA biogenesis and function. Mol Cell. doi:10.1016/j.molcel.2013.11.014

Yunis JJ, Yasmineh WG (1971) Heterochromatin, satellite DNA, and cell function. Structural DNA of eucaryotes may support and protect genes and aid in speciation. Science 174:1200–1209

Zaratiegui M, Castel SE, Irvine DV et al (2011) RNAi promotes heterochromatic silencing through replication-coupled release of RNA Pol II. Nature 479:135–138. doi:10.1038/nature10501

Zhang K, Fischer T, Porter RL et al (2011) Clr4/Suv39 and RNA quality control factors cooperate to trigger RNAi and suppress antisense RNA. Science 331:1624–1627. doi:10.1126/science.1198712

Zhang Z, Chang S-S, Zhang Z et al (2013) Homologous recombination as a mechanism to recognize repetitive DNA sequences in an RNAi pathway. Genes Dev 27:145–150. doi:10.1101/gad.209494.112

Zofall M, Yamanaka S, Reyes-Turcu FE et al (2012) RNA elimination machinery targeting meiotic mRNAs promotes facultative heterochromatin formation. Science 335:96–100. doi:10.1126/science.1211651

Chapter 14
Approaches for Dissecting RNA-Binding Protein Networks

Ana M. Matia-González and André P. Gerber

Abstract RNA-binding proteins (RBPs) play pivotal roles in post-transcriptional regulation, acting as master controllers of the RNA's life. The development of a variety of global analysis tools during the last decade has tremendously enriched our understanding of the function of RBPs and their implications in environmental and cellular cues. In this chapter, we discuss the properties and functions of the RBPs in fungi, and we outline the genome-wide approaches that have been developed during the last decade to systematically identify their RNA targets. We further sketch recent approaches aimed to search for novel RBPs, revealing that many more proteins such as metabolic enzymes could have additional functions in RNA regulation. Based on a rich dataset obtained in budding yeast, we further refer to attempts to integrate RNA targets for more than 60 RBPs into molecular networks to reveal structures and motifs. Finally, we elaborate on the future tasks and challenges on our way to complete our understanding of this highly connected and interwoven protein-RNA interaction network and its importance for cell physiology.

Contents

A. M. Matia-González · A. P. Gerber (✉)
Faculty of Health and Medical Sciences, Department of Microbial and Cellular Sciences, School of Biosciences and Medicine, University of Surrey, 17AY04 Stag Hill Campus, Guildford GU2 7XH, UK
e-mail: a.gerber@surrey.ac.uk

A. M. Matia-González
e-mail: a.matia@surrey.ac.uk

A. Sesma and T. von der Haar (eds.), *Fungal RNA Biology*,
DOI: 10.1007/978-3-319-05687-6_14,

Introduction

Gene expression is coordinated at different levels, which are highly interconnected and regulated (Maniatis and Reed 2002). Nevertheless, research on gene expression has been mainly focused on the first steps of this program, namely the transcriptional control mediated by transcription factors (TF) that activate genes by binding to specific DNA promoter sequences and recruit RNA polymerases for RNA synthesis. Whereas the roles of TFs, chromatin structure and its modifications are undisputed, it is now also becoming increasingly recognized that control of the later post-transcriptional steps has substantial regulatory impact with pivotal roles for development, metabolism, neuronal function, and aging (Costa-Mattioli et al. 2009; Tavernarakis 2008). Strong evidence for importance of post-transcriptional control is also given from quantitative proteomic studies showing that cellular mRNA transcript levels are limited indicators of protein abundance and thus, additional layers exist to control the coordination of protein synthesis (Gygi et al. 1999; Lu et al. 2007; Schwanhausser et al. 2011). Moreover, the spatial separation of transcription and translation by the nuclear membrane in eukaryotes demands structures to coordinate production and distribution of the large number of mRNA molecules present in a cell ($\sim$15,000 mRNA molecules in budding yeast).

Post-transcriptional coordination is mediated by a plethora of RNA-binding proteins (RBPs) that dynamically associate with specific subpopulations of transcripts. RBPs bind to RNAs forming dynamical functional units called ribonucleoprotein complexes (RNPs) that guide RNAs through RNA processing, maturation, transport, translation until finally mRNA decay (Moore 2005). A substantial fraction of RBPs have conserved orthologs and thus, the study of yeast RBPs may give valuable information on the function of human homologs, which may, or may not, be linked to human disease, e.g. (Scherrer et al. 2011). To date, more than 30 genetic human diseases—above all neurodegenerative disorders and cancers—have been linked to mutations in genes coding for RBPs (Lukong et al. 2008; Cooper et al. 2009; Castello et al. 2013). Furthermore, the understanding of the function and targets of essential RBPs in pathogenic fungi (e.g. *Candida albicans*) may become beneficial for the development of new drugs to protect us from infections.

Table 14.1 Examples of RNA-binding domains in budding yeast

Domain	Characteristics	RNA target	Proteins (*S. cerevisiae*)	Function
RRM	80–90 amino acids. ßαßßαß fold, four-stranded antiparallel ß-sheet packed against two α-helices	ssRNA	Hsh49, Cwc2	Pre-mRNA Splicing
			Nop6, Nop12	Ribsosome biogenesis
			Tif3, Sgn1	Translation initiation
			Hrp1, Pab1	Polyadenylation/ export/ translation initiation
DEAD box	Core sequence: DEAD (asp-glu-ala-asp)	ssRNA	Sub2p, Prp25p, Prp8p	Pre-mRNA Splicing
			Has1p	Ribosome biogenesis
			eIF4A	Translation initiation
KH	70 amino acids. Three stranded ß-sheet packed against three α-helices. Core sequence: (I/L/V)IGXXGXX(I/L/V)	ssRNA (4 nts) ssDNA	Khd1/Scp160	mRNA localization Translational control
PUF	8 repeats each consisting of α-helices	UGUR motif	Puf1	Localization
			Puf3	Localization, mRNA degradation
Zinc finger	Appropiately spaced cysteine-His_6	ssRNA	Gis2	Translation initiation/ degradation
			Nab2	mRNA Export
dsRBD	70 amino acids. αßßßα profile. Region with conserved positions including basic (R, K) and hydrophobic amino acids	dsRNA	Rnt1	rRNA processing, ribosome biogenesis
S1	70 amino acids organized in five-stranded antiparallel ß-barrel combined with three-helix residues *per turn*	ssRNA, dsRNA	Dis3, Rrp4	Exosome complex
			Rrp5	Ribosome assembly

This chapter provides an overview of RBPs and the approaches to identify their targets and networks in fungi. Most examples refer to the budding yeast *Saccharomyces cerevisiae* and the fission yeast *Schizosaccharomyces pombe*, as these are the most extensively studied fungi for post-transcriptional gene regulation.

Roles of RNA-Binding Proteins in Post-Transcriptional Gene Regulation

The Composition of RNA-Binding Proteins

RBPs consist of one or multiple small domains of about 35–90 amino acids that bind specifically to sequences or structural elements in the RNA (Table 14.1) (Auweter et al. 2006; Glisovic et al. 2008; Lunde et al. 2007). A compilation of annotated RBDs from Pfam/Interpro/Prosite databases reveals more than 160 experimentally validated RBDs (Emma Laing and Andre Gerber, unpublished results) (Anantharaman et al. 2002). Some very representative domains in yeast as well as in "higher eukaryotes" are the RNA recognition motifs (RRM; also called RBD or RNP domain), different types of Zn-finger domains, the S1 domain, and the DEAD-box helicase domain. The hnRNP K homology (KH) domain, the double-stranded RNA-binding domain (dsRBD), and the Pumilio/FBF (PUF or Pum-HD) domain are other well-characterized domains present in many regulatory RBPs although less prominent in yeast. Certain domains are quite unique and restricted to certain fungi, for example, the Piwi/Argonaute/Zwille (PAZ) is present in fission yeast *S. pombe* but absent in the budding yeast *S. cerevisiae*, which is in agreement with the lack of small-interfering RNAs (siRNA) related pathways in budding yeast (Anantharaman et al. 2002).

RBPs can contain one RBD, which typically binds 2–6 ribonucleotides. However, multiple copies of an RBD enables the recognition of larger and more complex RNA targets and enhances the specificity and affinity of binding (Auweter et al. 2006; Glisovic et al. 2008; Lunde et al. 2007). For example, the KH domain is found in large arrays in proteins collectively known as vigilins. Scp160p from *S. cerevisiae* contains 14 KH domains and the sequence homolog Vgl1 in fission yeast bears 11 KH domains. However, similar domains do not necessarily reflect similar functions. Scp160p is associated with polyribosomes and regulates translation (Lang and Fridovich-Keil 2000), whereas Vgl1 is involved in the transcriptional regulation during thermal stress response (Wen et al. 2010). Notably, RBDs can also be combined with other domains such helicases and enzymatic domains to make RNA helicases and RNA modification enzymes, respectively.

Bioinformatics analyses based on annotated RBDs suggested that eukaryotic genomes encode hundreds of RBPs (Anantharaman et al. 2002). In budding yeast, more than 550 proteins ($\sim$8 % of protein coding genes) are predicted or known to have functions in RNA-metabolism (Hogan et al. 2008). Thus, the content of RBPs is similar or even above than other important functional classes such as DNA-binding proteins or kinases, further reflecting the ancient importance of post-transcriptional gene regulation for cell function.

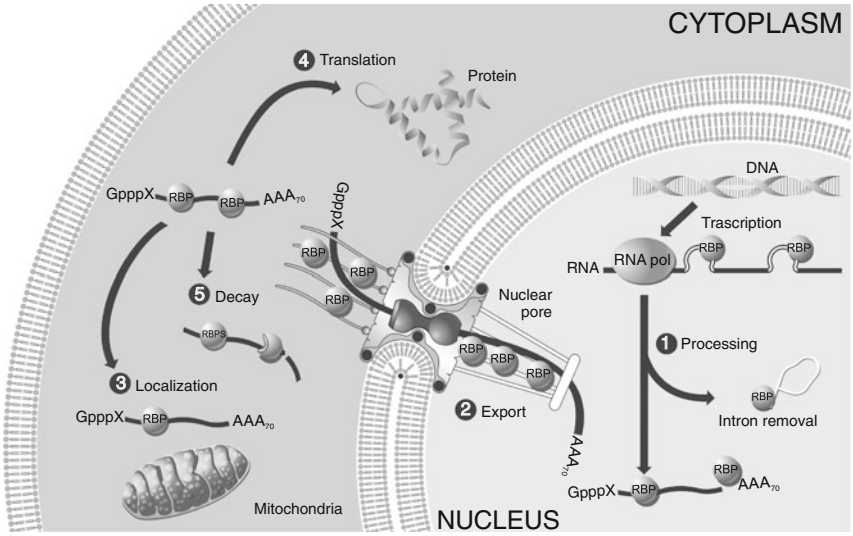


Fig. 14.1 Roles of RBPs in the post-transcriptional control of gene expression. RBPs mediate RNA processing reactions that involve splicing, capping and polyadenylation in the nucleus (*1*). The mRNAs are exported (*2*) through the nuclear pore complex and may be localized to specific subcellular locations (*3*) for translation (*4*). Eventually, RNAs are degraded by one of several decay pathways (*5*)

Functions of RNA-Binding Proteins

Upon synthesis of RNA precursors by one of the three RNA polymerases present in eukaryotic cells, the "early" RNA molecules are immediately bound by a host of RBPs forming RNPs. Depending on the nature of the primary transcript, which can be either a mRNA-precursor (pre-mRNA) or a non-coding RNAs (ncRNAs) such as ribosomal RNA (rRNA), transfer RNA (tRNA) and others, it undergoes a series of further processing and maturation steps to become fully effective. Focusing on pre-mRNAs, the transcription of these is accompanied by a series of processing reactions occurring in the nucleus: this involves capping, where a trimethyl guanosine modified cap is added at the 5′-end of messages; splicing, where introns are removed by a large RNP termed "spliceosome," and 3′-end cleavage followed by the addition of long polyadenylic acid (poly[A]) tail, which is important to maintain the mRNA stable and to enhance translation (Moore 2005) (Fig. 14.1; see Chap. 2 for splicing and Chap. 3 for 3′-end processing).

Once mRNA has been processed in the nucleus, it has to be exported to the cytoplasm for translation. The export occurs in three step process: the generation of an mRNA-carrier complex in the nucleus is followed by the translocation of the complex through the nuclear pore complex, where then the mRNA is released into the cytoplasm and carrier is recycled (Niño et al. 2013). Amongst several RBPs

involved in this process, Mex67p is central for mRNA export in budding yeast (Segref et al. 1997). Mex67p interacts with a large cohort of polyadenylated RNAs and tethers them to nucleoporin Nup85p, which is part of the nuclear pore complex, for the subsequent export through the pore (Niño et al. 2013; see Chap. 4).

Messenger RNAs may further undergo localization to specific subcellular regions by complexes consisting of motor proteins and RBPs or by the signal recognition particle (St Johnston 2005). A prominent example is the directed localization of *ASH1* and more than 20 other mRNAs to the bud-tip during cell division in the yeast *S. cerevisiae* (Heym and Niessing 2012; Shepard et al. 2003). In this case, the mRNAs are transported along actin cables by incorporation in a complex consisting of She2p, an RBP that directly interacts with the mRNA, Myo4p, which encodes a myosin and She3p, which serves as adaptor between She2p and Myo4p (Heym and Niessing 2012). During transport, mRNAs are bound by at least one translational repressor, such as Khd1p, that is modified and released at the destination site allowing the translation of mRNA cargo (Irie et al. 2002; Hasegawa et al. 2008; Paquin et al. 2007; see Chap. 6).

In happily growing yeast cells, most mRNAs ($\sim$80 % in yeast) are immediately assembled with multiple ribosomes, forming polysomes, to template protein synthesis. However, under non-favorable growth-condition i.e. environmental stress or starvation, mRNAs are removed from polysomes and either stored in stress granules or degraded in processing bodies (P-bodies) (Decker and Parker 2012). Numerous regulatory proteins target the initiation step of translation as translational regulation provides a rapid mechanism to control gene expression. For example Ded1p, a DEAD-box ATP-dependent RNA helicase, plays a pivotal role in the reversible storage of mRNAs. It is essential for translation initiation of all mRNAs but conversely, Ded1p can also inhibit translation through interaction with eIF4G in its ATP-bound status and accumulating bound mRNAs in stress granules (Hilliker et al. 2011). Eventually, those mRNAs that should not be longer available in the cells are degraded. In yeast, the major pathway for mRNA decay is initiated by deadenylation followed by decapping and 5′–3′ exonucleolytic digestion of the mRNA (Parker 2012; see Chap. 7). In addition, more selective decay pathways such as nonsense-mediated decay (NMD) eliminate mRNAs that contain premature translation initiation codons (PTCs) (Schoenberg and Maquat 2012; see Chap. 8). One example of an RBP that links translation with decay is Dhh1p, a DEAD-box RNA helicase that interacts with the 40S ribosomal subunit and thereby represses further movement of ribosomes along the mRNA. Such stalling eventually leads to decapping and degradation of the mRNA (Sweet et al. 2012).

Whereas some RBPs have very defined functions in the post-transcriptional regulation, numerous RBPs are involved in multiple steps of RNA regulation, providing strong connections between the different steps of gene expression program. These so-called hubs have generally more RNA targets (Mittal et al. 2011). For example, Pab1p, a major polyadenylate binding protein in yeast coordinates mRNA decay (Simon and Seraphin 2007), mRNA export (Brune et al. 2005) and translation initiation (Sachs and Davis 1989).

Approaches to Study RNP Complexes

Identification of RBP Targets

Studies of RBP-RNA interactions have historically relied on the identification of target transcripts bound by individual RBPs. *In vitro* selection of RNA sequences that bind RBPs with high affinity by systematic evolution of ligands by exponential enrichment (SELEX) can identify primary sequence recognition elements (Manley 2013; Riordan et al. 2011). However, the primary sequence elements identified in those screens are generally short and degenerated and thus often appear very frequently in the transcriptome complicating its use for *in silico* target description. An *in vivo* approach to screen or validate protein-RNA interactions is the yeast-three-hybrid (Y3H) assay (Bernstein et al. 2002; Martin 2012). By expressing chimeric proteins and RNA molecules in *S. cerevisiae*, this method allows monitoring of RNA-protein interactions by measuring the expression levels of a reporter gene (Martin 2012). Thereby, specific RNA targets can be used to characterize unknown RBPs and conversely, previously described RBPs can be used as bait to select new RNA targets. Although this method has been used in many instances in different fungi (Long et al. 2000; Yosefzon et al. 2011; Konig et al. 2009), its applicability to screen for RNA targets is strictly limited by library constraints and false positives (Martin 2012).

To add more selectivity for defining biologically relevant *in vivo* targets, a number of high-throughput "genomics" approaches have been developed that enable a fairly good description of the target spectra of RBPs. The genomic approach has also been successful to map the target RNA recognition elements (RREs) either directly and/or by inferring bioinformatics searches among selected RNA sequences. In the following, we summarize the methods applied to fungi to explore the RNA-protein interactions at a genome-wide level.

RNA-Binding Protein Immunoprecipitation-Microarray (RIP-Chip). This method involves the affinity purification of tagged RNP complexes under native conditions from cellular extracts (Fig. 14.2). Alternatively, antibodies that selectively recognize an epitope of a constituent protein can be used to immunopurify RNP complexes (Galgano and Gerber 2011). The purified RNPs are then dissociated into proteins and RNA, and the identity of the associated RNAs is determined on a global-scale with DNA microarrays or sequencing. Due to its simplicity and robustness, RIP-Chip has been widely used to uncover the RNA targets for more than 70 RBPs from *S. cerevisiae* (Hieronymus and Silver 2003; Gerber et al. 2004; Inada and Guthrie 2004; Hogan et al. 2008; Scherrer et al. 2010; Tsvetanova et al. 2010; Scherrer et al. 2011; Schenk et al. 2012), several RBPs from fission yeast *S. pombe* (Amorim and Mata 2009; Amorim et al. 2010; Matia-Gonzalez et al. 2013), and Khd4 from the filamentous fungus *Ustilago maydis* (Vollmeister et al. 2009).

Critical for RIP is the maintenance of RBP-RNA interactions during cell lysis and purification. Hence, the buffer components and the method to prepare the cell

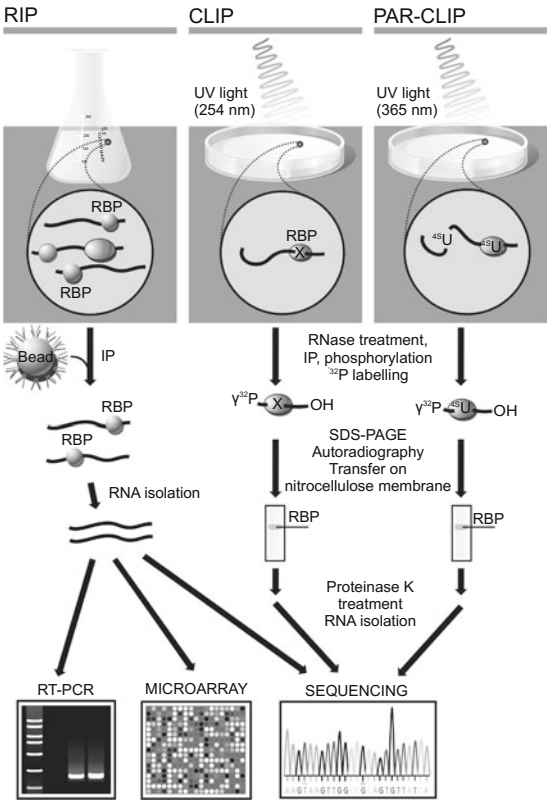


Fig. 14.2 Current methods to identify the RNA targets of RBPs. RIP captures RBPs from extracts with antibodies or via an epitope-tag under native conditions and bound RNAs are co-purified and analyzed after reverse transcription with qPCR, DNA microarrays or by sequencing. CLIP introduces UV irradiation of cells at 254 nm to crosslink all RNA-protein interactions. RNAs are mildly digested with RNase and RBP immunopurified. The bound RNA fragments are radioactively labelled and RNP complexes are separated by SDS-PAGE and transfer to a nitrocellulose membrane. RBP is digested with proteinase K and RNAs analyzed by deep sequencing. PAR-CLIP is a variation of CLIP that uses PAR analogs that are crosslinked to proteins under UV irradiation at 365 nm

lysates are critical to keep RNP complexes in a native state (Lopez de Heredia and Jansen 2004). Affinity isolation of RBPs and bound RNAs from the extracts is either performed with antibodies that selectively recognize an epitope of a constituent protein or by affinity purification of epitope-tagged proteins. The latter is very popular and to our knowledge, all RIP-Chip experiments performed in fungi have been performed with tagged versions of proteins that can be easily generated through genetic manipulation. The commonly used epitope in yeast is the tandem affinity purification (TAP) tag which contains a calmodulin binding peptide (CBP), followed by tobacco-etch virus protease (TEV protease) cleavage site and two IgG-binding domains of protein A (Rigaut et al. 1999).

To affinity-capture the RBP, porous agarose/sepharose beads are commonly used to immobilize antibodies or immunoglobulin G (IgG), due to low nonspecific absorption and excellent flow properties. However, magnetic beads are now gradually replacing the agarose/sepharose beads, enabling fast capture of the protein, short washings and efficient recovery as beads can be easily separated from the cell lysates with a magnet. After capturing on beads, several controlled washes are performed to remove low-affinity binders and contaminants and finally, the RBP and associated RNA are eluted from beads. Thereby, application of site-specific proteases such as TEV for cleaving-off the IgG-binding domains of the TAP-tag and releases the RBP from beads constitutes an additional purification step. Finally, RNA is purified from the eluate either by extraction with phenol/chloroform and alcohol or using commercial kits. To control for nonspecifically enriched RNAs, the same procedure is performed with wild-type cells lacking an affinity tag (Gerber et al. 2004), cells expressing an unrelated tagged-protein (Schenk et al. 2012) or a matrix that is non-coupled with antibodies (Galgano and Gerber 2011). Importantly, the purified RNP complexes can also be further subjected to mass-spectrometry (MS) analysis to identify directly the interacting proteins (=RNA independent), or RNA-dependent interacting proteins, giving an estimate for concurrent interacting proteins on the RNA targets for a particular RBP (Klass et al. 2013).

After RNA isolation, the pool of associated RNAs has been most often analyzed with DNA microarrays (RIP-Chip). In so doing, total RNA isolated from extracts (input) and from the immunopurified (IP) samples are reverse transcribed into cDNA which is further labelled with Cy3 and Cy5 fluorescent dyes, respectively. The labelled and occasionally amplified cDNA samples are then mixed and competitively hybridized to DNA microarrays. In this assay, the ratio of the two RNA populations at a given array element reflects the enrichment of the respective RNA by the affinity purification. Depending on the array format, different statistical analysis can be conducted to construct a profile of RNAs that are associated with the RBP under study. In our laboratory, we commonly used the approach outlined above by applying the dual colour array format and we further performed unpaired two classes significance analysis of microarrays (SAM) on median centered arrays to determine false discovery rates (FDRs) for each arrayed element.

The use of microarrays for studying RBP targets is limited by the representative DNA probes on the array. This limitation can avoided with high-throughput sequencing of RNAs (RNA-seq; see Chap. 15), which is a quantitative sequencing-based method for mapping transcribed regions, where complementary DNA fragments are analyzed by high-throughput sequencing and mapped to a reference genome or transcriptome (Nagalakshmi et al. 2008). Application of RNA-seq thus allows identification of all RNA sequences that co-purified with an RBP during RIP. Such high-throughput sequencing was introduced to analyze Argonaute-associated small RNAs in fission yeast (Buhler et al. 2008).

A major advantage of RIP is that it is a straightforward protocol that allows for the concomitant identification of RNA and protein components of RNPs.

Drawbacks of RIP are concerns that during the procedure certain RNAs or proteins may fall-off and others associate with RNP complexes, a concern nevertheless that applies to all kind of biochemical purification procedures. For the isolation of unstable RNP complexes, modifications of the RIP procedure and/or crosslinking either by UV or with chemicals prior to RIP-Chip can be amended to the protocol (San Paolo et al. 2009).

Crosslinking and Immunoprecipitation (CLIP). The UV-light crosslinking and immunoprecipitation (CLIP) protocol, was originally developed by Ule, Darnell and colleagues to map the interactions sites of Nova, an RBP that directs alternative pre-mRNA splicing in the brain (Ule et al. 2003; Konig et al. 2009). CLIP involves crosslinking of RNA-protein interactions with UV-light, followed by purification of the RBP with antibodies and mild RNA digestion to obtain a footprint of the RBP binding site on the RNA (Fig. 14.2). Thus, in principle it allows for direct mapping of binding sites and straightforward identification of RREs. Furthermore, CLIP is suitable for the detection of more transient and weak interactions. Combined with high-throughput sequencing, such HITS-CLIP has been adapted for the yeast *S. cerevisiae* to analyze targets and binding sites for Khd1p and Prp8p (Wolf et al. 2010; Li et al. 2013), and Rrm4, an RBP involved in the formation of polar-growing hyphae of *U. maydis* required for infection of plants (Konig et al. 2009).

CLIP is based on UV crosslinking of living cells or tissue samples at a wavelength of 254 nm that promotes irreversible covalent links between bases of the RNA and amino acids that are in close proximity (Greenberg 1979). Of note, chemicals, such as formaldehyde, can also be used for RNA-protein crosslinking (Moller et al. 1977) but formaldehyde also cross-links proteins to proteins and it may not enter the cores of large complexes. Hence, crosslinking with UV light is now commonly used to fix protein-RNA interactions. After crosslinking, cells are lysed and RNA is partially digested to an approximate length of 30–50 nucleotides with RNase. Of note, the RNase treatment is a crucial step of this protocol; the final result can be modified by either incomplete or extensive RNase treatment, but also the type of RNase used for the ribonuclease digestion may affect the sites identified (Kishore et al. 2011). Since protein-RNA binding sites are irreversibly crosslinked, relatively stringent wash conditions can be applied during IP, which can be performed analogous to RIP with antibodies or by capture of epitope-tagged RBPs. After IP, the crosslinked and bound RNA fragments are dephosphorylated, radioactively labelled at the 5′-end, and an RNA adaptor is ligated at the 3′-end. The RNP complex is further purified through denaturing gel electrophoresis and transferred to a nitrocellulose membrane to separate and remove non-covalently bound RNAs. After transfer to nitrocellulose, the radioactively labelled RNAs are used to pinpoint the excision of the protein-RNA complex from the membrane. The protein is then removed from the RNA by digestion with proteinase K, and an adaptor is ligated at the 5′-end of the RNA. The RNA is reverse transcribed and cDNAs are amplified by PCR using DNA primers with sequence complimentary to

the RNA linkers, generating a cDNA library which can be subjected to high-throughput sequencing (Konig et al. 2009).

A modification of CLIP, termed crosslinking and analysis of cDNAs (CRAC) has been specifically designed for the analysis protein-RNA interactions in larger RNPs containing many different proteins (Granneman et al. 2009). The method is based on yeast strains that were constructed to express RBPs fused to a modified TAP-tag, where the CBP is replaced with a fragment encoding six histidine residues (His_6). Purification of crosslinked RNP can so be complemented with an additional affinity purification step of the His-tagged RBP via nickel resin under denaturing conditions. This method has been used to map interaction sites of RPBs for budding yeast to highly structured ncRNAs, such as Rrp9 on U3 small nucleolar RNAs (snoRNAs) (Granneman et al. 2009), Rat1p on rRNA (Granneman et al. 2011); and to determine the *in vivo* binding sites for the nuclear RNA surveillance factors, Nrd1, Nab3, and the Trf4/5-Air1/2-Mtr4 complex (Wlotzka et al. 2011). Recently, CRAC was executed to identify the transcriptome-wide targets of 13 RNA processing, export and turnover factors in budding yeast (Tuck and Tollervey 2013). Of note, CLIP and CRAC provided similar data analyzing the *in vivo* RNA binding sites of Prp8p (Li et al. 2013).

Although the CLIP/CRAC method provides valuable information by identifying the direct binding sites on the RNAs it also has drawbacks. For instance, the UV irradiation imposes a stress to cells that will immediately activate stress response pathways (i.e. DNA damage response) which likely has direct impact on gene expression. UV exposure can also chemically and physically alter the RNP structures and cause some sequence bias due to the unequal photo reactivity between bases and amino acids (Gaillard and Aguilera 2008). Finally, the method is an elaborate multistep procedure that needs a skilled researcher to optimize the procedure.

Photoactivatable Ribonucleoside Enhanced Crosslinking and Immunopurification (PAR-CLIP). PAR-CLIP increases the efficiency of crosslinking, using photoactivatable ribonucleoside (PAR) analogs which are up-taken by cells with no toxic effect (Melvin et al. 1978). Briefly, cells are grown in the presence of 4-thiouracil (4sU) which is incorporated into RNA (Friedel and Dolken 2009). UV irradiation at 365 nm results in high-efficiency covalent crosslinking of RNA to protein (Hafner et al. 2010a, b). The RNP under study is then purified under native (if antibodies are applied) or optionally, under denaturing conditions for epitope-tagged proteins. After trimming unbound RNA with RNase, the RBP-protected fragments are ligated to linkers, converted to cDNA and subjected to high-throughput sequencing. Since the reverse transcription of the crosslinked 4sU leads to the mis-incorporation of guanosines on the opposite strand of RNA during reverse transcription, PAR-CLIP enables identification of binding sites in the RNA at single nucleotide resolution, which is represented by T to C transitions in the sequence (Hafner et al. 2010a). PAR-CLIP has been applied to map the binding sites of Nrd1, Nab3 and Sen1 which are all components of the yeast non-poly(A) termination pathway (Creamer et al. 2011; Jamonnak et al. 2011). The studies revealed novel ncRNA targets and confirmed previously known Nrd1 and Nab3

binding motifs in the RNA. Of note, the distribution of the RNA targets, e.g., ncRNAs, rRNAs, and tRNAs was similar to data obtained with CRAC (Wlotzka et al. 2011), suggesting that CRAC and PAR-CLIP have similar performance.

The PAR-CLIP approach has recently been extended to a global level and is referred to as global PAR-CLIP (gPAR-CLIP) (Freeberg et al. 2013). Instead of limiting the analysis to one particular RBP, the PAR-CLIP technique has been adopted to map all RBP binding sites across yeast non-translating mRNAs in different environmental conditions. Therefore, additional biochemical steps were implemented to capture RNA regions bound by the proteome: this includes sucrose density fractionation to capture only ribosome free RNPs, oligo-[dT] selection of RNAs to deplete abundant structural ncRNAs, and chemical biotinylation followed by capturing all bound proteins to RNA with streptavidin (Freeberg et al. 2013). After trimming unbound RNA with RNase, the RBP-protected fragments were ligated to linkers, converted to cDNA and subjected to Illumina sequencing. This comprehensive identification of RBP-RNA crosslinked sites enabled to derive further insight into the general properties of the RBP-RNA *in vivo*. For instance, it confirmed the overall accepted notion that regulatory RBPs preferentially interact with regions in 3′UTRs of targets. Moreover, comparative genomic analysis revealed that RBP crosslinking sites are generally conserved and that secondary structural elements in the RNA are constrained by protein binding. Interestingly, more than one third of crosslinked 3′UTR sites showed changes in RBP occupancy upon glucose or nitrogen deprivation, providing further evidence for dynamic control of mRNA fates upon stress.

Identification of RBPs

Canonical RBPs contain at least one of the more than 100 classified RBDs, allowing prediction of hundreds of RBPs in all kingdoms of life. However, it has been recognized for a while that also unconventional RBPs, which bind mRNAs through unique non-canonical RBDs appear to mediate essential RNA regulatory functions. Moreover, it was found that proteins with other established functions in cells such as metabolic enzymes in mammals can bind RNA and contribute to post-transcriptional gene regulation (Ciesla 2006; Castello et al. 2013). Evidence that RNA regulation through enzymes could also occur in fungi came from transcriptome-wide analysis of the RNA targets for two metabolic enzymes in the yeast *S. cerevisiae*, which reproducibly bound cellular RNAs (Hogan et al. 2008). Thus, the amount of RBPs in genomes could still be underestimated and the suspicion that many more proteins with RNA-binding function could exist in eukaryotes has motivated the development of new systematic experimental approaches to identify RBPs (Fig. 14.3).

Screen for RBPs with Protein Microarrays. First studies applied high-density protein microarrays to screen for proteins that interact with RNAs (Tsvetanova et al. 2010; Scherrer et al. 2010). Thereby, protein microarrays containing more

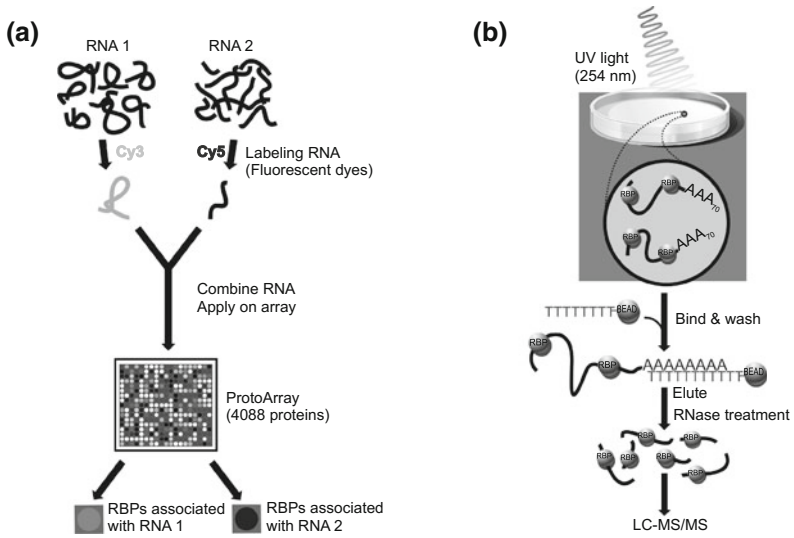


Fig. 14.3 **a** Protein microarrays that contain more than 4,000 yeast proteins (~70 % of the proteome), spotted in duplicates onto a modified glass slide are probed with a mixture containing specific RNAs labelled with fluorescent dyes. The arrays are then washed, scanned and analyzed for specific enrichment of labelled RNA. **b** Poly(A) mRNA pull-down. After UV crosslinking at 254 nm, the RNP complexes are isolated from the total RNA using oligo-[dT] beads. RNAs are digested and proteins identified via LC-MS/MS

than 4,000 proteins, each of them individually expressed and purified (representing 70 % of the yeast's proteome), were probed with fluorescently labelled RNAs, representing different types of RNA (e.g. total RNA vs. polyadenylated RNAs). After washing and scanning of the array, particular proteins interacting with RNAs are identified based on the fluorescent signals from interacting RNAs (Fig. 14.3a). Both studies recovered a substantial fraction of previously known RBPs, and at the same time identified dozens to hundreds of proteins that were not previously known to act as RBPs such as metabolic enzymes. For instance, in our study (Scherrer et al. 2010), we identified almost 200 proteins with more than 40 % of them having RNA-unrelated catalytic activities. Out of these, we selected 13 proteins, which have not been reported to bind RNA before, and identified their associated RNAs with RIP-Chip. We found significant association of most proteins with specific RNA target sets, many of them encoding proteins sharing similar functions or locations in the cell.

Despite its success in screening for new RBPs, protein microarrays have substantial limitations, some of them shared with other *in vitro* selection methods (i.e. SELEX). It is likely that not all of the proteins printed on the array are fully active and likewise, not all of the fluorescently labelled RNAs may reside in native conformation or interact with other RNAs in the pool. Nevertheless, the studies

defined many new RBPs and elucidation of their biological role will be an important task for future investigations.

Identification of Proteins Bound to RNA with Mass-Spectrometry. Another approach to identify the RBP proteome involves the purification of mRNAs followed by MS analysis of bound proteins. Previously applied to catch RBPs binding *in vitro* (Butter et al. 2009), Pat Brown and colleagues first introduced a concept which involved the purification of endogenously expressed polyadenylated mRNAs via oligo-[dT] columns. The mRNAs are subsequently degraded with RNase and proteins analyzed with liquid chromatography—tandem mass spectrometry (LC-MS/MS) (Fig. 14.3b). In that way, several potentially new RBPs, such as enzymes and components of COPI and COPII particles involved in vesicular transport and secretory pathways, could be identified as RNA-binders (Tsvetanova et al. 2010). Whereas this original study did not introduce any crosslinking to covalently fix protein-RNA interactions, the introduction of UV crosslinking combined with high performance quantitative MS tremendously improved the sensitivity of the assay, identifying hundreds of known and potentially new RBPs in mammalian cells (Castello et al. 2012; Baltz et al. 2012) and in the yeast *S. cerevisiae* (Mitchell et al. 2013).

In analogy to results obtained with protein microarrays, the MS based proteomic studies revealed that many proteins could have dual functions—e.g. connecting metabolism with RNA regulation—allowing direct connections between different layers of cellular control. Thus, the number of proteins that interact with RNAs is much higher than originally thought and could likely outperform related classes of proteins such as DNA-binding proteins or kinases. It suggests that proteins binding to RNA form a highly dense and robust post-transcriptional scaffold that effectively coordinates gene expression to ensure the integrity and stability of a cell's fate.

RNA-Protein Interaction Networks

Many features of RBP mediated gene regulation closely resemble those of TF: while TFs generally bind DNA motifs upstream of a given gene, RBPs bind to sequence or structural features of mRNA molecules which are often located in their untranslated regions. TFs are organized into transcription initiation complexes or "enhanceosomes," whereas RBPs assemble into highly dynamic transient RNPs that in a combinatorial may eventually define unique fates for individual target RNAs. Finally, both TFs and RBPs appear to bind targets that code for functionally and cytotopically related proteins.

Among fungi, the most comprehensive protein-interaction network has been obtained in the yeast *S. cerevisiae*. To date, the targets for more than 70 RBPs have been identified by RIP-Chip and *in vivo* binding sites for at least 15 RBPs have been mapped with CLIP/CRAC/PAR-CLIP. The data can now be used to construct RNA-protein interaction networks or post-transcriptional regulatory networks

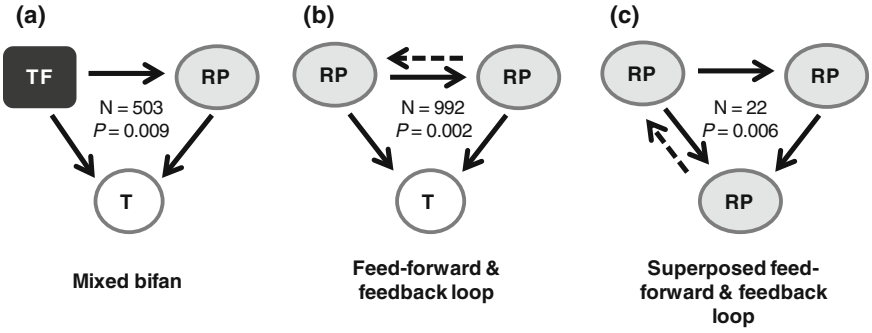


Fig. 14.4 Motifs overrepresented in post-transciptional networks. **a** In a mixed bifan motif, a TF and RBP jointly regulate one or more target genes. **b** In the feed-forward feedback loop (FFL-FBL) motif, RBPs tend to regulate common set of target RNAs and interact/regulate each other post-transcriptionally. **c** Superposed FFL-FBL motifs are overrepresented in the post-transcriptional network of RBP-RBP interactions. N refers to the number of occurrence and *P*-values indicate the significance of overrepresentation compared to a random control network (Joshi et al. 2012; Mittal et al. 2011)

Table 14.2 Properties of interaction networks

Parameter	TF-DNA	RBP-RNA	RBP-RBP
Regulator nodes	210	69	69
Number of nodes	3489	6161	69
Interactions	10,419	24,932	351
Network diameter	15	7	6
Clustering coefficient	0.029	0.197	0.203
Average number of neighbors	5.960	8.081	9.072
Self-loops (auto-regulation)	16	26	26
Characteristic path length	5.113	2.862	2.721

(Mittal et al. 2011; Joshi et al. 2011, 2012). Commonly, gene regulatory networks (GRNs) are represented in the form of a directional network, where regulators and targets are nodes that are interconnected by particular regulatory events (edges) that activate or repress the expression of the target (Fig. 14.4). To construct a first RBP-RNA network, the interactions of 69 RBPs with RNA determined by RIP-Chip in the yeast *S. cerevisiae* have been assembled (Mittal et al. 2011). The data can be imported into open-source software such as Cytoscape for network visualization and analysis (Saito et al. 2012). The so-constructed RBP-RNA network bears a total of 24,932 connections (edges) between 6,162 nodes (including the 69 regulator nodes) (Table 14.2). More than 80 % of the RNAs present in the transcriptome, which is based on annotated open reading frames (ORFs) and ncRNAs annotated in *Saccharomyces* genome database (SGD), have significant association with at least one of these 69 RBPs. Strikingly, the RBP-RNA interaction network has more interactions and is more dense than the related transcriptional network, which can be constructed from data obtained from a comprehensive chromatin

immunoprecipitation-CHIP (ChIP-CHIP) analysis that includes 210 TFs binding to DNA in promoter regions of genes, which we further refer to as the TF-DNA network (Harbison et al. 2004) (Table 14.2). Extrapolating the number of interactions among the 69 RBPs to the number of TF-DNA interactions analyzed, the RBP-RNA interaction network could contain about 8 times the number of interactions seen in the TF-DNA network (total ~80,500 interactions for 210 RBPs). The diameter of the posttranscriptional gene regulation network (which refers to the longest of all the shortest paths between pairs of nodes) is 7, indicating that two nodes of this network are separated by no more than seven edges, which is substantially shorter compared to the TF network where the diameter is 15. Further indication for a dense network is expressed by a higher average number of neighbors (~8 in the RBP-RNA *vs.* 6 in TF-DNA networks), a higher clustering coefficient, which expresses the trend for modular organization, and a shorter path lengths (an indication for the average number of edges to connect to nodes) (Table 14.2). In conclusion, these numbers provide evidence for an extensive and dense regulatory network at the post-transcriptional level.

The construction of GNRs allows the search for potential regulatory network motifs that are overrepresented (Alon 2007). The analysis can focus on one level (e.g. searching for motifs within one network) or on connections between different layered networks e.g. TF-DNA, RNA-RBP, or even RNA–RNA networks. Based on the comprehensive networks outlined above, a search for mixed motifs e.g. where a TF and RBP jointly regulate one or more targets indicates that the number of TF-RBP pairs is significantly lower than expected by chance (Joshi et al. 2012). Thus, RBPs do not generally partner with TFs to regulate common sets of target genes but rather selectively cooperate. However, an extended mixed bifan motif that implements an additional regulatory interaction between a TF and RBP in a feed-forward loop (FFL) is significantly enriched in yeast (Joshi et al. 2012) (Fig. 14.4a). Interestingly, this configuration, wherein an RBP is transcriptionally co-regulated with the targets of a TF could potentially acts as an effective noise buffer: fluctuations in TF levels will affect both target mRNA synthesis and post-transcriptional regulation by the RBPs such that altered production of the mRNA could be countered by corresponding changes of translation/decay rates. Such noise buffers are particularly effective if the RBP regulates the TF in a feedback loop. This occurs in the yeast network but the motif is not significantly overrepresented (Joshi et al. 2012).

Further considering the RBP-RNA network, it is often seen that two mutually regulating RBPs share a subset of target genes. In other words, many RBPs share substantial overlap with mRNA targets suggesting redundancy and effective combinatorial control (Hogan et al. 2008; Kanitz and Gerber 2009). But do RBPs also tend to regulate the expression of other RBPs at the post-transcriptional level? To address this question, the RBP-RBP post-transcriptional network can be extracted from the entire RBP-RNA network, thereby focusing on 69 RBPs and the 352 interactions among them (Table 14.2) (Mittal et al. 2011). Analysis of this network provides though further evidence for the "regulator of regulator" concept stating that RBPs have the tendency to regulate themselves or other RBPs at the

post-transcriptional level. First, one key finding is that autoregulatory loops are of much higher incidence among RBPs (26 out of 69 RBPs; 38 %) than among TFs (16 out of 210 TFs; 8 %) (Table 14.2) (Kanitz and Gerber 2009; Mittal et al. 2011; Joshi et al. 2011). Since feedback loops, in particular negative feedback loops imposed by repressor activities, are prototypical noise suppressing motifs, it could provide an explanation for the observation that autoregulatory RBPs tend to exhibit lower protein expression noise than non-autoregulatory ones (Mittal et al. 2011). Moreover, the set of potentially autoregulatory RBPs tend to be better connected with other RBPs, either by direct interactions or by binding to mRNAs coding for other RBPs (Mittal et al. 2011). Second, several feedback loops among the regulatory RBPs are overrepresented in yeast and likely in other organisms as well (Joshi et al. 2012; Mittal et al. 2011) (Fig. 14.4b). This includes FFL motifs which are also prominent in TF network motifs, wherein a top-level RBP controls two target transcripts and one of these further regulates the other target (Fig 14.4b). Again, feedback loops are prototypical noise suppressing motifs in regulatory networks (Alon 2007), and thus could be an effective tool to minimize noise of key regulators of downstream cellular processes. The same FFL is also overrepresented in the RBP-RBP network, where it is found 269 times with a p-value of <0.005. Furthermore, there are 22 instances of a superposed FFL-FBL motif in the RBP-RBP network, where an RBP X controls the target RBPs Y and Z, with Y controlling Z, and Z in turn controlling X (Fig. 14.4c). Of note, if such reciprocal interactions occur between X and Z nodes, each one could control the other and thereby change the directionality of the FFL, resulting in different regulatory output on downstream targets.

In conclusion, RBPs show a strong tendency to control each other building on strong autoregulatory and control circuits that could lead to low expression noise. Moreover, the "regulator of regulator" concept may not be restricted to fungi but rather be a universal principle for RBP networks in all organisms (Pullmann et al. 2007).

Conclusions

To date, transcriptome-wide identification of RNA targets has been conducted for more than 80 RBPs from the budding yeast *S. cerevisiae* and other fungi. Therefore, a variety of experimental approaches have been developed based on the purification of an RBP or an epitope-tagged version thereof using antibodies or high-affinity ligands, followed by the identification of bound RNAs—which are crosslinked to the protein or not—with DNA microarrays or high-throughput sequencing. Besides the importance of knowing the targets of RBPs and their binding sites on the RNA, which provides valuable information for further investigations on the biological function and physiological impact for cell physiology, the global studies have also given substantial insight into underlying concepts of RNA biology at a systems level. In particular, we wish to point out that the

data obtained from these global studies during the last decade strongly support and further extended the "post-transcriptional operon model" initially proposed by Jack Keene and colleagues (Keene and Tenenbaum 2002). In analogy to prokaryotic operons, the model predicted that RBPs coordinate groups of mRNAs coding for functionally related proteins. *Cis-acting* elements in the mRNA may provide the means to mimic the coordinated regulatory advantages of clustering genes into polycistronic operons (Keene and Tenenbaum 2002; Keene 2007). In the following, we briefly summarize some of the major insights obtained from the global studies explained in above sections.

First, RBPs bind to unique sets of RNA. Thereby, the number of the associated RNAs can vary widely revealing unique associations with 20 to >1,000 distinct transcripts *per* RBP. Of note, the number of targets for RBPs may be five to ten times higher than for TFs (Table 14.2). Second, sequence or structural elements in the RNAs are enriched among targets defining the binding site for the RBP (Gerber et al. 2004; Hogan et al. 2008; Scherrer et al. 2011). Thereby, it has been confirmed that regulatory RBPs preferentially target sequences in the 3′UTR of mRNAs, at least for non-translating mRNAs (Freeberg et al. 2013). Third, bound mRNAs preferentially encode functionally and/or cytotopically related proteins. This is perhaps best exemplified with the yeast Puf proteins, the targets of which share striking common cytotopic features (Gerber et al. 2004). Fourth, the spectra of targets for RBPs strongly overlap with targets of other RBPs, suggesting strong combinatorial binding of RBPs (Hogan et al. 2008). Such combinatorial regulation may greatly impinge on the regulatory potential of the RBPs—breaking-up simple linear correlations between RBPs and the fates of its targets. For example, the yeast RBP Khd1 oppositely controls the expression of mRNA targets: some messages are increasingly expressed ("up-regulated") whereas others are repressed ("downregulated") by Khd1p, which is likely due to combinatorial control with other RBPs (Hasegawa et al. 2008). Fifth, the RNA-protein interaction network is very dynamic and responds to environmental or developmental signals by altering the RNA and protein content of RNPs (Freeberg et al. 2013). Thereby, post-translational modification of the RBP (e.g. phosphorylation) can alter subcellular localization or RNA-binding activity of the RBP (e.g. Paquin et al. 2007). Sixth, RBPs tend to regulate themselves (auto-regulation) or other regulatory proteins such as RBPs and TF. First network analyses of post-transcriptional interactions between RBPs suggest that it forms a dense and intertwined network, with potential for about 40 % of the RBPs to auto-regulate transcript levels. Finally, based on network analysis, it can be speculated that the RPB-RNA network acts as noise buffer for gene expression in eukaryotes (Joshi et al. 2011; Mittal et al. 2011). At the end, the RBP-RNA network, whether linked to noise control or not, likely contributes to the more coherent responses upon cellular perturbation. This may be reflected by the fact that the translatome (referring to the entirety of mRNAs associated with ribosomes for translation) shows more coherent responses compared to the transcriptome upon different sorts of stress (Halbeisen and Gerber 2009; Joshi et al. 2011), emphasizing the importance for translational control in cell physiology and development.

Perspectives

Ribonomic analysis revealed that RBPs preferentially associate with messages that share common functional and structural attributes suggesting the presence of a highly complex and interweaved post-transcriptional regulatory system. In addition, unravelling the RNA targets for particular RBPs has led to insights into their molecular and physiological function (Schenk et al. 2012; Rowe 2014). Nevertheless, it is still largely unknown how perturbation of RNA regulons and the underlying RNA-protein interactions are linked to stress, development and disease.

Yeast is a formidable tool to study systematically the dynamics of mRNP and RNA regulon composition in response to environmental perturbations or differentiation. Moreover, the integration of global RNA-protein interaction data with measurements of steady-state mRNA levels, ribosome/polysome profiling data and large-scale quantitative proteomics data will become and even more popular approach to determine the impact of particular post-transcriptional regulators on splicing, translation and/or mRNA degradation. Whatsoever, the meaningful integration of data obtained from different levels of the gene expression program will remain a major challenge and need further development of intuitive computational analysis tools that are readily available to support biologists in the generation of new hypotheses. In this context, strong collaborative efforts between biologists and computational scientist will be required to construct and interpret network models in systems biology.

In the long term, based on the detection and quantification of network motifs in gene regulatory systems, we will certainly make advances in the development of predictive models. We shall then be keen to adapt knowledge obtained in yeast or other model organisms to the human system to infer potentially evolutionary conserved circuits. The circuits may well be linked to human disease, further justifying their dissection in simple models such as yeast where intrinsic or extrinsic perturbations can be easily accomplished. As seen many times in the past, the back and forth between yeast and humans may generate a positive forward loop that promotes understanding of complex malfunctions such as cancer and neurodegenerative disease.

Finally, yeast or other fungi could be used for the development of novel drugs targeting RNA regulons either through specific interactions with *cis*-acting elements or through direct targeting of regulatory RBPs. Although the targeting of RBPs by small molecules has been less favored in the past as considered less likely for success, there have been recently some promising attempts (Meisner et al. 2007). Conversely, the application of novel types of antisense molecules to block RNA-binding sites are making progress and several clinical trials are ongoing (Southwell et al. 2012). The growing recognition for importance and impact of RNA biology in general and the RBP network in particular, will likely promote the development of alternative approaches to develop new cures for human disease and add benefit for human well-being.

Acknowledgements We are grateful to Jose Luis González Molero for preparing figures and members of the Gerber lab for discussion. AM is supported by a Sinergia grant (CRSII3_141942) from the Swiss National Science Foundation to APG.

References

Alon U (2007) Network motifs: theory and experimental approaches. Nat Rev Genet 8:450–461

Amorim MJ, Cotobal C, Duncan C, Mata J (2010) Global coordination of transcriptional control and mRNA decay during cellular differentiation. Mol Syst Biol 6:380

Amorim MJ, Mata J (2009) Rng3, a member of the UCS family of myosin co-chaperones, associates with myosin heavy chains cotranslationally. EMBO Rep 10:186–191

Anantharaman V, Koonin EV, Aravind L (2002) Comparative genomics and evolution of proteins involved in RNA metabolism. Nucleic Acids Res 30:1427–1464

Auweter SD, Oberstrass FC, Allain FH (2006) Sequence-specific binding of single-stranded RNA: is there a code for recognition? Nucleic Acids Res 34:4943–4959

Baltz AG, Munschauer M, Schwanhausser B, Vasile A, Murakawa Y, Schueler M, Youngs N, Penfold-Brown D, Drew K, Milek M, Wyler E, Bonneau R, Selbach M, Dieterich C, Landthaler M (2012) The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol Cell 46:674–690

Bernstein DS, Buter N, Stumpf C, Wickens M (2002) Analyzing mRNA-protein complexes using a yeast three-hybrid system. Methods 26:123–141

Brune C, Munchel SE, Fischer N, Podtelejnikov AV, Weis K (2005) Yeast poly(A)-binding protein Pab1 shuttles between the nucleus and the cytoplasm and functions in mRNA export. RNA 11:517–531

Buhler M, Spies N, Bartel DP, Moazed D (2008) TRAMP-mediated RNA surveillance prevents spurious entry of RNAs into the *Schizosaccharomyces pombe* siRNA pathway. Nat Struct Mol Biol 15:1015–1023

Butter F, Scheibe M, Morl M, Mann M (2009) Unbiased RNA-protein interaction screen by quantitative proteomics. Proc Natl Acad Sci U S A 106:10626–10631

Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, Krijgsveld J, Hentze MW (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell 149:1393–1406

Castello A, Fischer B, Hentze MW, Preiss T (2013) RNA-binding proteins in Mendelian disease. Trends Genet 29:318–327

Ciesla J (2006) Metabolic enzymes that bind RNA: yet another level of cellular regulatory network? Acta Biochim Pol 53:11–32

Cooper TA, Wan L, Dreyfuss G (2009) RNA and disease. Cell 136:777–793

Costa-Mattioli M, Sossin WS, Klann E, Sonenberg N (2009) Translational control of long-lasting synaptic plasticity and memory. Neuron 61:10–26

Creamer TJ, Darby MM, Jamonnak N, Schaughency P, Hao H, Wheelan SJ, Corden JL (2011) Transcriptome-wide binding sites for components of the *Saccharomyces cerevisiae* non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. PLoS Genet 7:e1002329

Decker CJ, Parker R (2012) P-bodies and stress granules: possible roles in the control of translation and mRNA degradation. Cold Spring Harb Perspect Biol 4:a012286

Freeberg MA, Han T, Moresco JJ, Kong A, Yang YC, Lu ZJ, Yates JR, Kim JK (2013) Pervasive and dynamic protein binding sites of the mRNA transcriptome in *Saccharomyces cerevisiae*. Genome Biol 14:R13

Friedel CC, Dolken L (2009) Metabolic tagging and purification of nascent RNA: implications for transcriptomics. Mol BioSyst 5:1271–1278

Gaillard H, Aguilera A (2008) A novel class of mRNA-containing cytoplasmic granules are produced in response to UV-irradiation. Mol Biol Cell 19:4980–4992

Galgano A, Gerber AP (2011) RNA-binding protein immunopurification-microarray (RIP-Chip) analysis to profile localized RNAs. Methods Mol Biol 714:369–385

Gerber AP, Herschlag D, Brown PO (2004) Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. PLoS Biol 2:E79

Glisovic T, Bachorik JL, Yong J, Dreyfuss G (2008) RNA-binding proteins and post-transcriptional gene regulation. FEBS Lett 582:1977–1986

Granneman S, Kudla G, Petfalski E, Tollervey D (2009) Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. Proc Natl Acad Sci USA 106:9613–9618

Granneman S, Petfalski E, Tollervey D (2011) A cluster of ribosome synthesis factors regulate pre-rRNA folding and 5.8S rRNA maturation by the Rat1 exonuclease. EMBO J 30:4006–4019

Greenberg JR (1979) Ultraviolet light-induced crosslinking of mRNA to proteins. Nucleic Acids Res 6:715–732

Gygi SP, Rochon Y, Franza BR, Aebersold R (1999) Correlation between protein and mRNA abundance in yeast. Mol Cell Biol 19:1720–1730

Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr, Jungkamp AC, Munschauer M, Ulrich A, Wardle GS, Dewell S, Zavolan M, Tuschl T (2010a) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141:129–141

Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jungkamp AC, Munschauer M, Ulrich A, Wardle GS, Dewell S, Zavolan M, Tuschl T (2010b) PAR-CliP—a method to identify transcriptome-wide the binding sites of RNA binding proteins. J Vis Exp 41:pii2034

Halbeisen RE, Gerber AP (2009) Stress-Dependent Coordination of Transcriptome and Translatome in Yeast. PLoS Biol 7:e105

Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok DK, Kellis M, Rolfe PA, Takusagawa KT, Lander ES, Gifford DK, Fraenkel E, Young RA (2004) Transcriptional regulatory code of a eukaryotic genome. Nature 431:99–104

Hasegawa Y, Irie K, Gerber AP (2008) Distinct roles for Khd1p in the localization and expression of bud-localized mRNAs in yeast. RNA 14:2333–2347

Heym RG, Niessing D (2012) Principles of mRNA transport in yeast. Cell Mol Life Sci 69:1843–1853

Hieronymus H, Silver PA (2003) Genome-wide analysis of RNA-protein interactions illustrates specificity of the mRNA export machinery. Nat Genet 33:155–161

Hilliker A, Gao Z, Jankowsky E, Parker R (2011) The DEAD-box protein Ded1 modulates translation by the formation and resolution of an eIF4F-mRNA complex. Mol Cell 43:962–972

Hogan DJ, Riordan DP, Gerber AP, Herschlag D, Brown PO (2008) Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. PLoS Biol 6:e255

Inada M, Guthrie C (2004) Identification of Lhp1p-associated RNAs by microarray analysis in *Saccharomyces cerevisiae* reveals association with coding and noncoding RNAs. Proc Natl Acad Sci USA 101:434–439

Irie K, Tadauchi T, Takizawa PA, Vale RD, Matsumoto K, Herskowitz I (2002) The Khd1 protein, which has three KH RNA-binding motifs, is required for proper localization of *ASH1* mRNA in yeast. EMBO J 21:1158–1167

Jamonnak N, Creamer TJ, Darby MM, Schaughency P, Wheelan SJ, Corden JL (2011) Yeast Nrd1, Nab3, and Sen1 transcriptome-wide binding maps suggest multiple roles in post-transcriptional RNA processing. RNA 17:2011–2025

Joshi A, Beck Y, Michoel T (2012) Post-transcriptional regulatory networks play a key role in noise reduction that is conserved from micro-organisms to mammals. FEBS J 279:3501–3512

Joshi A, van de Peer Y, Michoel T (2011) Structural and functional organization of RNA regulons in the post-transcriptional regulatory network of yeast. Nucleic Acids Res 39:9108–9117

Kanitz A, Gerber AP (2009) Circuitry of mRNA regulation. WIRES Syst Biol Med 2:wsbm.55

Keene JD (2007) RNA regulons: coordination of post-transcriptional events. Nat Rev Genet 8:533–543

Keene JD, Tenenbaum SA (2002) Eukaryotic mRNPs may represent posttranscriptional operons. Mol Cell 9:1161–1167

Kishore S, Jaskiewicz L, Burger L, Hausser J, Khorshid M, Zavolan M (2011) A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. Nat Methods 8:559–564

Klass DM, Scheibe M, Butter F, Hogan GJ, Mann M, Brown PO (2013) Quantitative proteomic analysis reveals concurrent RNA-protein interactions and identifies new RNA-binding proteins in *Saccharomyces cerevisiae*. Genome Res 23:1028–1038

Konig J, Baumann S, Koepke J, Pohlmann T, Zarnack K, Feldbrugge M (2009) The fungal RNA-binding protein Rrm4 mediates long-distance transport of ubi1 and rho3 mRNAs. EMBO J 28:1855–1866

Lang BD, Fridovich-Keil JL (2000) Scp160p, a multiple KH-domain protein, is a component of mRNP complexes in yeast. Nucleic Acids Res 28:1576–1584

Li X, Zhang W, Xu T, Ramsey J, Zhang L, Hill R, Hansen KC, Hesselberth JR, Zhao R (2013) Comprehensive in vivo RNA-binding site analyses reveal a role of Prp8 in spliceosomal assembly. Nucleic Acids Res 41:3805–3818

Long RM, Gu W, Lorimer E, Singer RH, Chartrand P (2000) She2p is a novel RNA-binding protein that recruits the Myo4p-She3p complex to *ASH1* mRNA. EMBO J 19:6592–6601

Lopez De Heredia M, JANSEN RP (2004) RNA integrity as a quality indicator during the first steps of RNP purifications : a comparison of yeast lysis methods. BMC Biochem 5:14

Lu P, Vogel C, Wang R, Yao X, Marcotte EM (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nat Biotechnol 25:117–124

Lukong KE, Chang KW, Khandjian EW, Richard S (2008) RNA-binding proteins in human genetic disease. Trends Genet 24:416–425

Lunde BM, Moore C, Varani G (2007) RNA-binding proteins: modular design for efficient function. Nat Rev Mol Cell Biol 8:479–490

Maniatis T, Reed R (2002) An extensive network of coupling among gene expression machines. Nature 416:499–506

Manley JL (2013) SELEX to identify protein-binding sites on RNA. Cold Spring Harb Protoc 2013:156–163

Martin F (2012) Fifteen years of the yeast three-hybrid system: RNA-protein interactions under investigation. Methods 58:367–375

Matia-Gonzalez AM, Hasan A, Moe GH, Mata J, Rodriguez-Gabriel MA (2013) Functional characterization of Upf1 targets in *Schizosaccharomyces pombe*. RNA Biol 10:1057–1065

Meisner NC, Hintersteiner M, Mueller K, Bauer R, Seifert JM, Naegeli HU, Ottl J, Oberer L, Guenat C, Moss S, Harrer N, Woisetschlaeger M, Buehler C, Uhl V, Auer M (2007) Identification and mechanistic characterization of low-molecular-weight inhibitors for HuR. Nat Chem Biol 3:508–515

Melvin WT, Milne HB, Slater AA, Allen HJ, Keir HM (1978) Incorporation of 6-thioguanosine and 4-thiouridine into RNA. Application to isolation of newly synthesised RNA by affinity chromatography. Eur J Biochem 92:373–379

Mitchell SF, Jain S, She M, Parker R (2013) Global analysis of yeast mRNPs. Nat Struct Mol Biol 20:127–133

Mittal N, Scherrer T, Janga SC, Gerber AP (2011) Interplay between post-transcriptional and post-translational interactions of RNA-binding proteins. J Mol Biol 409:466–479

Moller K, Rinke J, Ross A, Buddle G, Brimacombe R (1977) The use of formaldehyde in RNA-protein cross-linking studies with ribosomal subunits from *Escherichia coli*. Eur J Biochem 76:175–187

Moore MJ (2005) From birth to death: the complex lives of eukaryotic mRNAs. Science 309:1514–1518

Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 320:1344–1349

Niño CA, Hérissant L, Babour A, Dargemont C (2013) mRNA nuclear export in yeast. Chem Rev 113:8523–8545

Paquin N, Menade M, Poirier G, Donato D, Drouet E, Chartrand P (2007) Local activation of yeast *ASH1* mRNA translation through phosphorylation of Khd1p by the casein kinase Yck1p. Mol Cell 26:795–809

Parker R (2012) RNA degradation in *Saccharomyces cerevisae*. Genetics 191:671–702

Pullmann R Jr, Kim HH, Abdelmohsen K, Lal A, Martindale JL, Yang X, Gorospe M (2007) Analysis of turnover and translation regulatory RNA-binding protein expression through binding to cognate mRNAs. Mol Cell Biol 27:6265–6278

Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B (1999) A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol 17:1030–1032

Riordan DP, Herschlag D, Brown PO (2011) Identification of RNA recognition elements in the *Saccharomyces cerevisiae* transcriptome. Nucleic Acids Res 39:1501–1509

Rowe W, Kershaw CJ, Castelli LM, Costello JL, Ashe MP, Grant CM, Sims PF, Pavitt GD, Hubbard SJ (2014) Puf3p induces translational repression of genes linked to oxidative stress. Nucleic Acids Res 42(21):1026-41

Sachs AB, Davis RW (1989) The poly(A) binding protein is required for poly(A) shortening and 60S ribosomal subunit-dependent translation initiation. Cell 58:857–867

Saito R, Smoot ME, Ono K, Ruscheinski J, Wang PL, Lotia S, Pico AR, Bader GD, Ideker T (2012) A travel guide to Cytoscape plugins. Nat Methods 9:1069–1076

San Paolo S, Vanacova S, Schenk L, Scherrer T, Blank D, Keller W, Gerber AP (2009) Distinct roles of non-canonical poly(A) polymerases in RNA metabolism. PLoS Genet 5:e1000555

Schenk L, Meinel DM, Strasser K, Gerber AP (2012) La-motif-dependent mRNA association with Slf1 promotes copper detoxification in yeast. RNA 18:449–461

Scherrer T, Femmer C, Schiess R, Aebersold R, Gerber AP (2011) Defining potentially conserved RNA regulons of homologous zinc-finger RNA-binding proteins. Genome Biol 12:R3

Scherrer T, Mittal N, Janga SC, Gerber AP (2010) A screen for RNA-binding proteins in yeast indicates dual functions for many enzymes. PLoS ONE 5:e15499

Schoenberg DR, Maquat LE (2012) Regulation of cytoplasmic mRNA decay. Nat Rev Genet 13:246–259

Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M (2011) Global quantification of mammalian gene expression control. Nature 473:337–342

Segref A, Sharma K, Doye V, Hellwig A, Huber J, Luhrmann R, Hurt E (1997) Mex67p, a novel factor for nuclear mRNA export, binds to both poly(A)$^+$ RNA and nuclear pores. EMBO J 16:3256–3271

Shepard KA, Gerber AP, Jambhekar A, Takizawa PA, Brown PO, Herschlag D, Derisi JL, Vale RD (2003) Widespread cytoplasmic mRNA transport in yeast: identification of 22 bud-localized transcripts using DNA microarray analysis. Proc Natl Acad Sci USA 100:11429–11434

Simon E, Seraphin B (2007) A specific role for the C-terminal region of the Poly(A)-binding protein in mRNA decay. Nucleic Acids Res 35:6017–6028

Southwell AL, Skotte NH, Bennett CF, Hayden MR (2012) Antisense oligonucleotide therapeutics for inherited neurodegenerative diseases. Trends Mol Med 18:634–643

St Johnston D (2005) Moving messages: the intracellular localization of mRNAs. Nat Rev Mol Cell Biol 6:363–375

Sweet T, Kovalak C, Coller J (2012) The DEAD-box protein Dhh1 promotes decapping by slowing ribosome movement. PLoS Biol 10:e1001342

Tavernarakis N (2008) Ageing and the regulation of protein synthesis: a balancing act? Trends Cell Biol 18:228–235

Tsvetanova NG, Klass DM, Salzman J, Brown PO (2010) Proteome-wide search reveals unexpected RNA-binding proteins in *Saccharomyces cerevisiae*. PLoS One 5:e12671

Tuck AC, Tollervey D (2013) A transcriptome-wide atlas of RNP composition reveals diverse classes of mRNAs and lncRNAs. Cell 154:996–1009

Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB (2003) CLIP identifies Nova-regulated RNA networks in the brain. Science 302:1212–1215

Vollmeister E, Haag C, Zarnack K, Baumann S, Konig J, Mannhaupt G, Feldbrugge M (2009) Tandem KH domains of Khd4 recognize AUACCC and are essential for regulation of morphology as well as pathogenicity in *Ustilago maydis*. RNA 15:2206–2218

Wen WL, Stevenson AL, Wang CY, Chen HJ, Kearsey SE, Norbury CJ, Watt S, Bahler J, Wang SW (2010) Vgl1, a multi-KH domain protein, is a novel component of the fission yeast stress granules required for cell survival under thermal stress. Nucleic Acids Res 38:6555–6566

Wlotzka W, Kudla G, Granneman S, Tollervey D (2011) The nuclear RNA polymerase II surveillance system targets polymerase III transcripts. EMBO J 30:1790–1803

Wolf JJ, Dowell RD, Mahony S, Rabani M, Gifford DK, Fink GR (2010) Feed-forward regulation of a cell fate determinant by an RNA-binding protein generates asymmetry in yeast. Genetics 185:513–522

Yosefzon Y, Koh YY, Chritton JJ, Lande A, Leibovich L, Barziv L, Petzold C, Yakhini Z, Mandel-Gutfreund Y, Wickens M, Arava Y (2011) Divergent RNA binding specificity of yeast Puf2p. RNA 17:1479–1488

Chapter 15
Bioinformatics Tools for Next-Generation RNA Sequencing Analysis

Marco Marconi, Julio Rodriguez-Romero, Ane Sesma and Mark D. Wilkinson

Abstract The purpose of this chapter is to introduce the reader to some of the most popular bioinformatics tools and resources available for RNA analysis. The introduction of RNA next-generation sequencing led to an explosion in the amount of quantitative transcript sequence data, which necessitated the development of adequate tools to process and make a sense of these rich and complex datasets. A large number of programs, platforms, and databases dedicated to RNA analysis have been produced over the past approximately 20 years; however, like so much other bioinformatics software, only a small portion of them are still available and in-use. As such, we will focus only on those tools and applications still in common use. This chapter is composed of three sections: the description of the general protocols for RNA sequence (generically called RNA-Seq) analyses, an outline of the most common approaches to map polyadenylation sites, and a brief introduction to noncoding RNA (ncRNA) analysis. The first section will describe the composition of steps within a typical RNA-Seq study: the experimental design, the sequencing methods, the data quality control, the read mapping, and the differential expression analysis. The second section will introduce a few recent methods developed to map polyadenylation sites: the experimental protocols (which are variations of RNA-Seq), polyadenylation site databases and prediction programs, and cis-regulatory elements discovery. The third and final section will present several of the ncRNA databases and prediction tools.

M. Marconi · J. Rodriguez-Romero · A. Sesma · M. D. Wilkinson (✉)
Centre for Plant Biotechnology and Genomics, Department of Biotechnology, Technical University of Madrid, Campus Montegancedo, AutopistaM-40 (Km.38), 28223 Pozuelo de Alarcón, Madrid, Spain
e-mail: markw@illuminae.com

A. Sesma and T. von der Haar (eds.), *Fungal RNA Biology*,
DOI: 10.1007/978-3-319-05687-6_15,

Contents

Whole Transcriptome Shotgun Sequencing (RNA-Seq)

One of the most common bioinformatics techniques applied to RNA is transcriptome analysis. The transcriptome is the complete complement and quantity of transcripts in a cell at a certain developmental stage or physiological condition. Compared to the relatively static genome, the transcriptome is highly variable and its constitution depends on many factors. Next-generation sequencing has facilitated the deep sequencing of RNA molecules allowing both highly accurate nucleotide coverage, as well as detection of low copy number transcripts, the analysis of differential expression, posttranscriptional changes, splicing, ncRNA population, polyadenylation sites, and gene fusions.

Before RNA-Seq the most common technique used to study transcriptome profiling was the expression microarray. Microarrays make use of nucleic acid hybridization with DNA probes matching to specific target sequences, based on a previously built cDNA library. A limitation of this approach is that it requires, and assumes, "prior knowledge" of the transcriptome—it is only possible to detect transcripts mapping to already-known transcribed regions that have been spotted on the array. Other limitations affecting microarray data are, the lack of sensitivity for very high/low-expression genes, weak concordance between platforms, no (straightforward) alternative splicing or gene-fusion detection, and no (straightforward) detection of allele-specific expression. The RNA-Seq approach has already overcome most or all of these disadvantages, and will be the primary focus of this chapter.

Experimental Design

Experimental design requires an initial determination of the experimental goals. For example, whether the objective is to perform a transcriptome assembly, analyze differential expression, or identify rare transcripts. The characteristics of the system are also important. Are the genomes large and complex? Are introns present? Is alternative-splicing a consideration? Is a reference genome available? Beyond these questions, the design itself can be straightforward, e.g., comparing two kinds of sample (the subject- mutant, drug treatment, etc. -vs. a control) or complex, using a variety of mutants, treatments, and time points. In the latter case, clearly, more sophisticated statistical analysis will be required.

Biological replicates are essential for differential expression analysis, while they are not required for transcriptome assembly. In RNA-Seq, technical replications are not usually needed, since there is very little technical variation between platforms. Other technical details to consider include if the sequencing will be single or pair-ended, the depth and length of the reads, and/or the pooling of samples.

Sequencing

Several commercial platforms exist for RNA sequencing, including Illumina sequencing by synthesis (GAIIx, HiSeq and MiSeq) (Kozarewa et al. 2009), SoliD (McKernan et al. 2009), and 454 pyrosequencing(Elahi and Ronaghi 2004). The choice of platform depends primarily on what kind of data is desired (Polyadenylated RNA, total RNA, small RNA…) and the cost (Mardis 2008; Shendure and Ji 2008).

The standard RNA-sequencing protocol begins with cDNA library preparation, consisting of several steps including RNA isolation, purification, and fragmentation (these first steps are strongly affected by what kind of RNA one wants to sequence), cDNA synthesis using random or oligo (dT) primers, adapter ligation, size selection, and PCR amplification.

With the Illumina platform, the cDNA to be sequenced is fragmented into about 200 bp segments which are end-ligated with adapters and hybridized to the flow cell. A localized PCR reaction is then performed on each strand to amplify it and create a cluster of the same molecule. The sequencing follows, adding fluorescent-tagged nucleotides to the cluster. Each nucleotide has the 3′ OH group blocked which ensures the incorporation of one nucleotide at a time. The light emitted by each nucleotide ligation is observed by the machine to determine the exact sequence. The Illumina read length is approximately 35 bases, but over one billion reads are generated on each run. The major error type is substitution, rather than deletion or insertion.

In the SOLID platform the cDNA library is similarly fragmented and end-ligated with adapters. The fragments are attached to small paramagnetic beads and

emulsion PCR is performed to amplify the fragments. In contrast with the other sequencing platforms, sequencing by synthesis is performed by utilizing DNA ligase, rather than polymerase. Each cycle of sequencing involves the ligation of a degenerate population of fluorescently labeled universal octamer primers, with a known set of di-nucleotide patterns at the 5′ end. The octamer is then cleaved, and another round of octamer ligation ensues. After several rounds, the ligated strand is melted off and replaced by a fresh adapter primer of n–1 length. Subsequent rounds of ligation are, therefore, offset by one versus the previous round. The pattern of fluorescence in these n/n–1 pairs of reads makes it possible to determine which di-nucleotide pattern matches the template. This approach involves the examination of the bases twice in a cycle, which decreases the error rates. SOLID read lengths are 25–35 bp, and each sequencing run yields approximately 1 billion reads.

Roche/454 FLX Pyrosequencer technology is based on the pyrosequencing method, which utilizes the enzymes ATP sulfurylase and luciferase. After the incorporation of each nucleotide by DNA polymerase, a pyrophosphate is released, which further takes part in downstream light-producing reactions. The amount of light is proportional to the incorporated number of nucleotides. The 454 FLX instrument generates ∼400,000 reads per instrument-run, and the reads are 200–400 bp. The greatest advantage of this platform is the read length, which is the longest of all next-generation technologies. Although sequencing on the 454 platform is more expensive than sequencing on the Illumina platform, it is the best choice for *de novo* transcriptome assembly and some meta genomics applications where assembly is part of the quantitation pipeline.

Data Quality Control

The typical output from a next-generation sequencing run is a FASTQ file (Cock et al. 2010). FASTQ is a file format utilizing the ASCII character set, with features similar to FASTA files, but with additional information about base-calling quality. These data are necessary to identify and remove contaminants and low quality sequences. A widely used tool to assess data quality is FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). FastQC allows visualization of various aspects related to the general quality of the sequencing run through its analysis of the input FASTQ file(s). Fast QC makes it possible to check, among other things, quality scores across bases or whole reads, GC distribution, adapter contamination, level of sequence duplication, and nonrandom hybridization of random primers. Depending on the results of this initial quality control analysis, further data processing may be required. For example, “data prepping” is needed when the format is somehow incorrect, the reads contain low quality bases, there is a contamination of adapter sequences or the data is out of synchronization. However, if none of these upstream problems are encountered, it is usually better not to perform any preprocessing of the data. There are many helpful tools

available on the Internet to process FASTQ libraries, the most popular are the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) and the ea-utils (https://code.google.com/p/ea-utils/).

Read Mapping

After the read library has been properly processed, it can then be aligned to the reference. Alignment algorithms for next-generation RNA sequencing data are somewhat more sophisticated than comparable DNA genomic aligners, since they have to take into account the presence of introns in the reference genome, which will split reads over multiple locations in the reference. This comes at a computational cost (i.e., speed), and therefore if splicing events are not expected in the experimental system being studied, it is faster to use unspliced aligners.

RNA-Seq aligners generally fall into two groups: Burrows Wheeler Transform (BWT) mappers, and Hash table mappers. BWT mappers are generally faster than hash table mappers but allow fewer mismatches and have limited indel detection. Popular BWT mappers are Top Hat (Trapnell et al. 2009) and Map Splice (Wang et al. 2010) for spliced alignments, and BWA (Li and Durbin 2010) and bowtie (Langmead and Salzberg 2012) for unspliced alignments. Popular hash table mappers include GSNAP (Wu and Nacu 2010) (spliced) or SHRiMP (David et al. 2011) and Stampy (Lunter and Goodson 2011) (unspliced).

The output of a transcriptome-wide mapping is generally given in the form of a SAM file (or its binary version, BAM). Sam files are tab-delimited text files that contain sequence alignment data. The sam toolkit provides tools for their manipulation and statistics reports (Li et al. 2009).

In case the reference genome is not fully annotated, it is often desirable to perform *de novo* discovery of novel transcripts. A popular tool for assembling sample reads and merging them into a single unified transcriptome dataset is the Cufflinks package (Trapnell et al. 2012).

Differential Expression

Characterizing the level of gene expression in a cell is a useful way to determine how the transcriptome is represented at, for example, a given stage of development, or in a specific type of tissue and/or in mutant strains. Although gene expression and protein expression are not always directly correlated (Arvas et al. 2011), due to other levels of regulation (RNAi interference, posttranscriptional modification, translational regulation, etc.), transcriptome profiling provides significant amounts of both quantitative and qualitative information about the state of the cell, and is therefore of great interest.

Where microarrays provide an "analog" signal of continuously variable levels of expression, RNA-Seq provides a more digital, discrete measurement, which follows a negative binomial distribution more suitable for mathematical and statistical analysis.

Cuffdiff from the Cufflinks package is a user-friendly differential expression program (Trapnell et al. 2012), which can be used to find significant changes in transcript expression, splicing, and promoter use through a pairwise comparison. It takes, as input, the result of an alignment as described earlier in SAM/BAM format, and the reference genome in GTF/GFF format. It is relatively easy to use and it is well-documented online.

Requiring more familiarity with statistics and the R programming language, the EdgeR (Robinson et al. 2010) and DeSEQ (Anders and Huber 2010) packages from the Bioconductor framework (Reimers and Carey 2006) could be considered, as they offer more power and flexibility than Cuffdiff. The input for these application interfaces is usually a raw gene/transcript read count, without the need of a reference genome (even if the reference genome is still necessary to calculate the read count) and they return the results in a tabular format, which can be saved as a regular text file, a comma separated values (CSV) file, or other spreadsheet formats.

Approaches to Mapping Polyadenylation Sites

Common structural elements and primary sequences are found within eukaryotic messenger RNAs (mRNAs). Cleavage and polyadenylation of the emerging pre-mRNA is necessary for maturation of almost all eukaryotic mRNAs (see Chap. 3). The 3′ untranslated regions (UTRs) of mRNAs, situated directly after the stop codon, harbor signals for transcript stability, localization, and translational control (Di Giammartino et al. 2011; Lutz and Moreira 2010). The control of mRNA expression by 3′UTRs is mediated by transacting elements, including RNA-binding proteins and micro RNAs (miRNAs), which interact with cis-regulatory elements within the 3′UTRs (Sandberg et al. 2008). The polyadenosine tail or poly(A), is crucial for the nuclear export, translation, and stability of mRNA. The tail is shortened over time, and, when it is short enough, the mRNA is enzymatically degraded (see Chap. 7).

The presence of multiple polyadenylation sites in the same gene leads to mRNA isoforms containing different coding sequences (CDS) and/or variable 3′UTRs (Tian et al. 2005). This phenomenon involves alternative cleavage and polyadenylation (APA) and can play a significant role in mRNA metabolism by controlling the length of 3′UTRs and the contained cis-regulatory elements, as described in more detail in Chap. 3 (Di Giammartino et al. 2011; Lutz and Moreira 2010). Dynamic regulation of 3′UTR by APA has been reported in different tissue types (Zhang et al. 2005; Wang et al. 2008), during cell proliferation, differentiation and development (Sandberg et al. 2008), in cancer cell transformation (Singh

et al. 2009; Flavell et al. 2008), and in response to extracellular stimuli (Flavell et al. 2008). The presence of polyadenylation sites within introns and upstream exons has not been fully researched at the genomic level. In addition, to what extent APA regulates long noncoding RNAs (lncRNAs), which are increasingly found to play important roles in the cell (Chen and Carmichael 2010; Wang and Chang 2011), is largely unknown.

Experimental Protocols

The growth in bioinformatics analysis of 3′UTR processing was prompted by the availability of large amounts of cDNA/EST sequences containing poly(A) tails in public databases like NCBI Gen Bank or dbEST (http://www.ncbi.nlm.nih.gov/dbEST/). Preliminary studies showed that a great proportion of genes have alternate polyadenylation, and that the cleavage process is largely imprecise, with different poly(A) sites having distinct nucleotide composition in surrounding genomic regions. With the emergence of next-generation sequencing technologies, it has become possible to sequence, in extreme depth, entire transcriptomes within days, making it possible to study this phenomenon in more detail.

Unfortunately, the standard protocol for mRNA sequencing (RNA-Seq) and its close variants are quite inefficient with respect to mapping poly(A) cut sites, since only a very small percentage of sequencing reads contain poly(A) tails; as such, most of the sequencing effort is wasted (Fig. 15.1). Recently, protocols have been developed to enrich transcript ends of poly(A)$^+$ mRNAs prior to RNA-Seq specifically to focus on the identification of poly(A) sites (Yoon and Brem 2010; Fox-Walsh et al. 2011; Ozsolak et al. 2010; Beck et al. 2010; Shepard et al. 2011; Jan et al. 2011; Derti et al. 2012; Morris et al. 2012).

Initial attempts to perform efficient poly(A) sites sequencing used oligo (dT) as primers for the sequencing reaction. This method ensured that only those fragments that contained a poly(A) tail were part of the amplified set (Shepard et al. 2011). These early methods suffered from several drawbacks. In protocols where sequencing proceeds from the middle of the transcript toward the poly(A) tail (Yoon and Brem 2010; Fox-Walsh et al. 2011; Ozsolak et al. 2010; Beck et al. 2010), the readout strongly depends on a stringent size selection to ensure that each read contains enough of the 3′UTR to map to the genome while still reaching the poly(A) tail. Illumina-based protocols where sequencing starts at the 3′-end and reads through the poly(A) site, encounter problems of low sequence quality (Shepard et al. 2011; Derti et al. 2012), apparently due to some type of "desynchronization" taking-place during base incorporation. It is known that polymerase slippage occurs when reading through long stretches of adenosines, and thus likely results in mis-priming of the sequencing oligo during clustering. A possible solution to this desynchronization is the shortening of the poly(A) tail but even a few remaining nucleotides of the poly(A) tail can compromise the delineation of

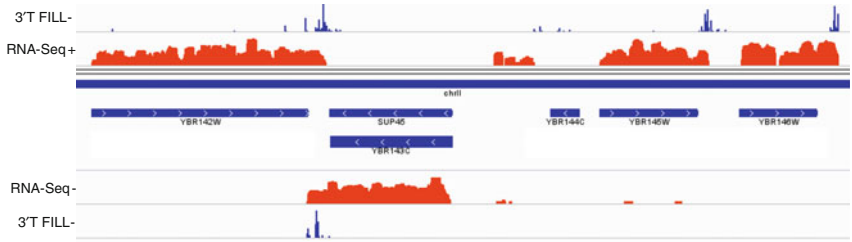


Fig. 15.1 Comparison of RNA-Seq and the 3′T-fillmethod for poly(A) site mapping (Wilkening et al. 2013). In RNA-Seq, only a small fraction of the signals arise from the 3′-ends of transcripts. With the 3′T-fill protocol, detection of poly(A) sites is more specific

clusters by the sequencing software (Jan et al. 2011). Moreover, performing the poly(A) shortening directly on the RNA could damage the fragments.

To overcome this problem, some poly(A) site-specific mapping protocols have been developed to avoid reading through the poly(A) tail (Wilkening et al. 2013; Hoque et al. 2013). In these protocols, the poly(A) stretch is filled in with dTTPs before the sequencing reaction starts. In this way, it is possible to help ensure that the sequencing reads begin immediately after the poly(A) tail, that is, on the exact nucleotide where the polyadenylation machinery performed the cut.

Another issue to be considered when mapping poly(A) sites is related to "internal priming." Identification of poly(A)s typically relies on the cDNA sequence corresponding to the poly(A) tail, which is generated by oligo(dT)-based reverse transcription (Brockman et al. 2005). However, oligo (dT) can also prime at internal adenosine-rich sequences anywhere in the genome, which as a result of this mismatched priming are completely converted to adenosine stretches in the final sequence, becoming indistinguishable from the sequence derived from the real poly(A) tail (Nam et al. 2002). A common way to overcome the internal priming issue is by eliminating putative poly(A) cut sites mapping to genomic A-rich regions from the final dataset, usually during the bioinformatics analysis. This approach, however, presents two potential problems. First, it is difficult to recognize and remove all the false positives accurately, and second, in converse, it also has the potential to misidentify and therefore discard real poly(A) sites. This problem is largely alleviated through the use of dTTP-filling approaches as described above.

Finally, some mRNAs have short oligo(A) tails that are synthesized by non-canonical poly(A) polymerases, such as those involved in exosome-mediated RNA decay (Schmidt and Norbury 2010). Overcoming this problem involves additional procedures, including handling the conditions during the RNA capture and purification that maximally distinguishes RNAs with long poly(A) tails (typical mRNAs) from those with short poly(A) tails. Briefly, after fragmentation of RNA, poly(A)-containing RNA fragments are captured onto magnetic beads coated with a chimeric oligonucleotide (oligo), which contained 45 thymidines (Ts) at the 5′ portion and 5′ uridines (Us) at the 3′ portion (Hoque et al. 2013).

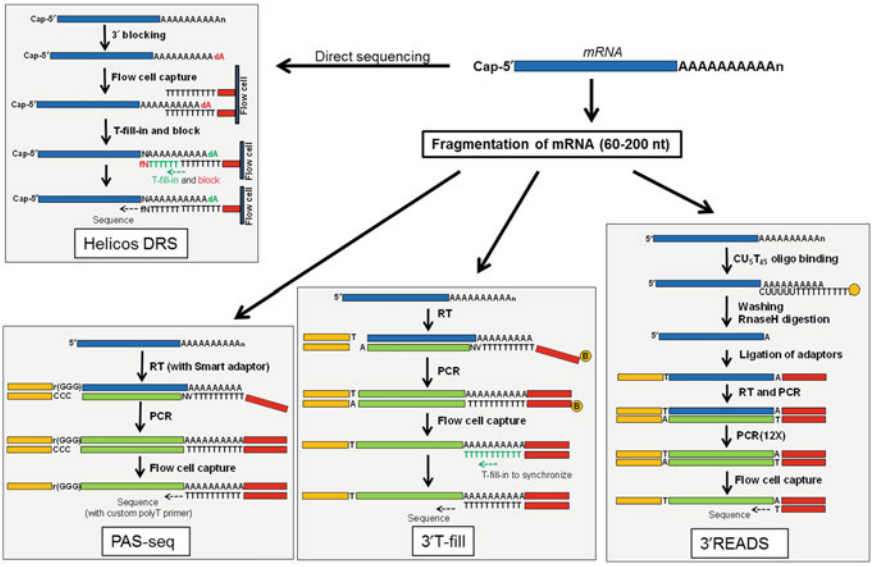


Fig. 15.2 Schematic comparison of the PAS-Seq (Shepard et al. 2011), 3′T-fill (Wilkening et al. 2013), DRS (Ozsolak et al. 2010) and 3′READS (Hoque et al. 2013) protocols

The methods described above require RNA to be converted to complementary DNA (cDNA) using reverse transcriptase prior to sequencing. This step itself has been shown to introduce many biases and artifacts. Contemporary approaches that minimize these errors include "single-molecule direct RNA sequencing (DRS)", where RNA molecules are sequenced directly in a massively parallel manner without the need of RNA conversion to cDNA, or other biasing sample manipulations such as ligation and amplification (Fig. 15.2). A technical comparison between all the methods described above is shown in Table 15.1.

Polyadenylation Site Databases and Prediction Tools

Polyadenylation has not been analyzed and mapped as well as other transcriptional processes like splicing (see Chap. 2), but some databases and tools are available for most of the model species.

Databases

PACdb is a database of mRNA 3′UTR processing sites (Brockman et al. 2005). PACdb relies on Expressed Sequence Tags (ESTs) to identify putative 3′UTR processing sites, which can be used to retrieve 3′UTR length and sequence based

Table 15.1 Comparison between some experimental poly(A) sequencing protocols

Study	Sequence start	Platform	Species	Percentage Poly(A) identifying reads (%)	Percentage of mapped reads in $3'UTR^3$ (%)	Limitations
RNA-Seq	RNA-Seq	HiSeq	*Saccharomyces cerevisiae*	0.18	19	Not designed for mapping poly(A) sites
Beck et al. 2010	5′ end	GAII	*Homo sapiens*	–	33–40	Stringent size selection required, PE reads do not add value
Ozsolak et al. 2010 (Helicos Drs)	3′ end	Helicos DRS	*S. cerevisiae* *H. sapiens*	–	91	Short read length, limited availability, lower throughput
Yoon and Brem 2010	5′ end	GAII	*S. cerevisiae*	–	35	Not strand-specific, 50 % of reads sequence through a poly(T) stretch resulting in low quality
Fox-Walsh et al. 2011	5′ end	GAII	*H. sapiens*	–	85	Stringent size selection required, PE reads do not add value
Shepard et al. 2011 (PAS-seq)	3′ end	GAII	*H. sapiens*	–	63	Decreased quality due to desynchronized sequencing start (polymerase slippage on cluster)
Jan et al. 2011	3′ end	GAII	*Caenorhabditis elegans*	–	–	Protocol includes many RNA manipulation steps, reduced cluster recognition due to remaining A's
Pelechano et al. 2012	5′ end	HiSeq	*S. cerevisiae*	47.00	89	Stringent size selection required, PE reads do not add value
Derti et al. 2012	3′ end	GAIIx	*Five mammals*	–	17–27	Decreased quality due to desynchronized sequencing start (polymerase slippage on cluster)
Wilkening et al. 2013 (3′Fill)	3′ end	HiSeq	*S. cerevisiae*	74.40	92	Access to cluster station required
Hoque et al. 2013 (3′Reads)	5′ end	GAIIx	*Mus musculus*	56	66.3–91.1	Enzymatic digestion with RnaseH, which adds an extra manipulation step

The studies denoted in brackets were introduced in this chapter

on other genomic datapoints. PACdb stores many known and putative 3′UTR processing sites from several organisms.

PolyA_DB 2 contains poly(A) sites identified for genes in several vertebrate species, including human, mouse, rat, chicken, and zebrafish, using alignments between cDNA/ESTs and genome sequences (Lee et al. 2007). This database also includes synthetic genome regions for human poly(A) sites in seven other vertebrates and cis-elements information adjacent to poly(A) sites. Trace sequences are used to provide additional evidence for poly(A/T) tails in cDNA/ESTs. The updated database is intended to broaden poly(A) site coverage in vertebrate genomes, and provides a means to assess the authenticity of poly(A) sites identified by bioinformatics tools.

UTRdb and UTRsite are curated databases of 5′ and 3′ untranslated sequences of eukaryotic mRNAs, derived from several sources of primary data (Grillo et al. 2010). Experimentally validated functional motifs are annotated in the UTR site database where specific information on the functional motifs and cross-links to interacting regulatory protein are provided. UTR entries have been organized in a gene-centric structure to better visualize and retrieve 5′ and 3′UTR variants generated by alternative initiation and termination of transcription and alternative splicing. Experimentally validated miRNA targets and conserved sequence elements are also annotated. The integration of UTRdb with genomic data has allowed the implementation of an efficient annotation system and a powerful retrieval resource for the selection and extraction of specific UTR subsets.

Poly(A) Site Prediction Tools

PolyA_SVM is an analysis and prediction tool for mRNA polyadenylation sites utilizing a Support Vector Machine (SVM)(Cheng et al. 2006). This program requires a file containing DNA/RNA sequences in the FASTA format as input, and (1) makes prediction for putative poly(A) sites and/or (2) generates an output dataset indicating the occurrences of various known cis-elements. The program is implemented in PERL and runs under UNIX/LINUX systems.

PolyApred is another SVM-based method for the prediction of polyadenylation signals in human DNA sequence (Ahmed et al. 2009). In this method SVM models have been developed for predicting poly(A) signals in a DNA sequence using 100 nucleotide supstream and downstream of this signal.

Dragon PolyA Spotter is a poly(A) motif prediction method based on properties of human genomic DNA sequence surrounding a poly(A) motif (Kalkatawi et al. 2012). These properties include thermodynamic, physicochemica, and statistical characteristics. For predictions, an Artificial Neural Network and Random Forrest models were developed. These models are trained to recognize 12 most common poly(A) motifs in human DNA.

Cis-Acting Regulatory Elements and Prediction Programs

During the 3′ end processing of pre-mRNAs, a correct cleavage site is identified by the polyadenylation machinery, which first cuts the pre-mRNA and then synthesizes the polyadenosine tail. This cleavage site is identified due to the presence of specific polyadenylation signals carried by the pre-mRNA itself. The first confirmed cleavage-associated sequence motif, AAUAAA, is usually located 10–30 bp upstream from the pre-mRNA cut site (Wilson et al. 1977). Subsequently, additional flanking signals (auxiliary elements) were shown to be required for both 3′-end cleavage and polyadenylation of pre-mRNAs as well as to promote downstream transcriptional termination (see Chap. 3 for more details).

Bioinformatics Tools for Motif Discovery

Several bioinformatics tools are available to discover conserved motifs in the regions surrounding polyadenylation sites. The most popular online suite for motif-based sequence analysis is the MEME Suite (Bailey et al. 2009). The MEME Suite allows several types of analyses:

- discovery of motifs on groups of related DNA or protein sequences
- search sequence databases for known motifs
- comparison of a motif to all motifs within a database of motifs
- associate identified motifs with Gene Ontology terms based on their putative target genes and motif enrichments.

Here, a walk-thorough example is provided of the use of the MEME suite to identify, de novo, a motif known to be important in locating the polyadenylation machinery. As a background to this example, it has long been known in *S. cerevisiae* and other yeasts, that there is an adenosine-rich efficiency element (EE) upstream the canonical AAUAAA signal (Mischo and Proudfoot 2013), which likely functions to enhance the affinity of the cleavage complex with the pre-mRNA, or to allow the recognition of "weak" polyadenylation signals that differ from the canonical. Polyadenylation-related proteins such as *Hrp1* have been shown to interact with the EE in yeast (Perez-Canadillas 2006). The most common described motif representing the EE is the TAYRTA hexamer. We will now attempt to discover this motif without a previous knowledge of its existence using data from a poly(A) sequencing (poly(A)-seq) experiment and bioinformatics tools from the MEME suite.

The poly(A)-seq data can be derived from any of the previously described methods, or even from EST libraries. Here, the *S. cerevisae* poly(A)-seq data from the 3′T-FILL protocol described previously will be used (Wilkening et al. 2013). After identifying the poly(A) site in each sequence and aligning them to the genome reference (see Reads mapping), a region of interest of 100 bp is selected for the MEME search. This task can be accomplished using the NCBI FASTA

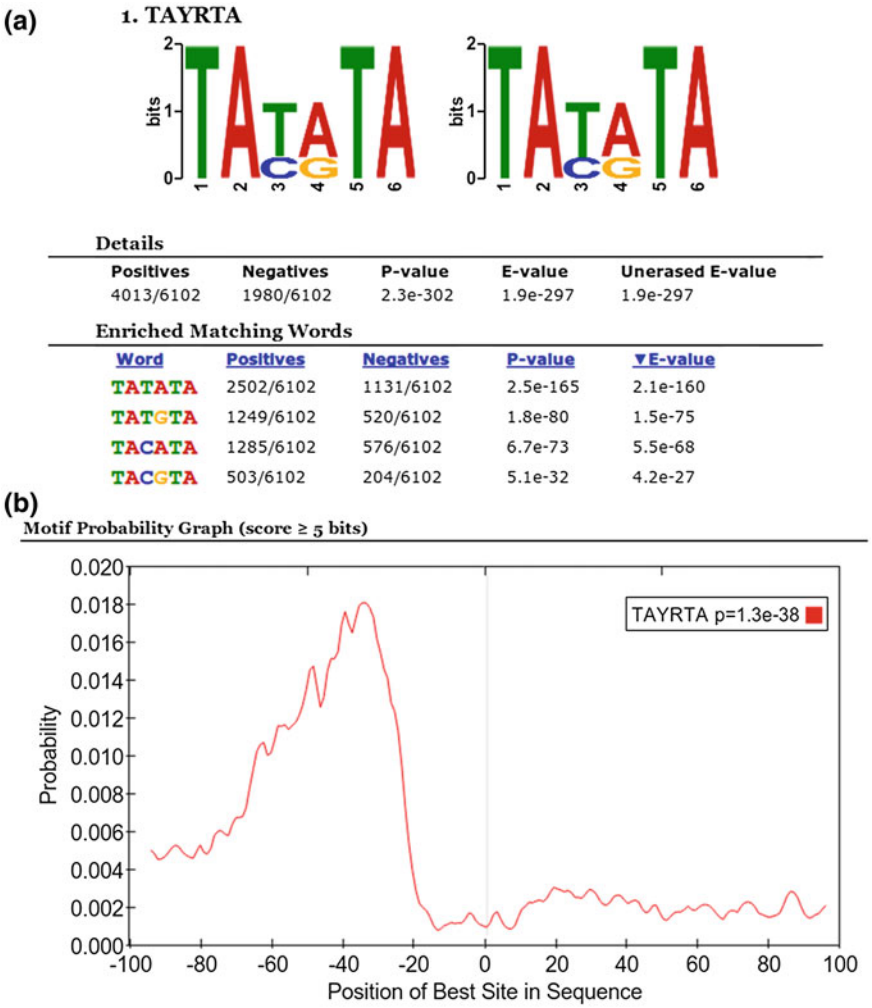


Fig. 15.3 Snapshots of the step-by-step example to discover the TAYRTA motif in the upstream region of *S. cerevisae* poly(A) sites using the MEME suite. **a** Result of the execution of DREME. **b** Plot of the TAYRTA motif using CENTRIMO

formatting tools like fastacmd, or programmatically by script with the aid of libraries like BioPerl (Stajich et al. 2002) or BioPython (Cock et al. 2009). Once a list of the 100 bp upstream region for every poly(A) site has been generated in the form of a FASTA file, the DREME tool included in the MEME suite, can be used to detect over-represented short motifs. The TAYRTA motif is the first to be detected with a fairly high E-value (Fig. 15.3a). In order to analyze the localization of the TAYRTA motif in the upstream region, the output from DREME can be passed to another MEME tool, CENTRIMO, which will show the local enrichment

of the newly discovered motif in the sequence set. For a more global visualization, CENTRIMO can be run over both the upstream and downstream region surrounding the poly(A) site, for example, from –100 to +100 bp (Fig. 15.3b).

Another wide collection of modular tools for the detection of cis-regulatory elements in genome sequences is RSAT (http://rsat.ulb.ac.be/rsat/) (Thomas-Chollier et al. 2011). This web site offers a series of tools dedicated to the detection of regulatory signals in noncoding sequences. The only input required is a list of genes of interest (e.g., a family of co-regulated genes), and the tool(s) will return the upstream sequences over a desired distance, putative regulatory signals discovered, the matching positions of these signals in the source dataset or in whole genomes, graphical display of the results in the form of a feature map (Shen et al. 2008). It also offers a series of facilities for random model generation and statistical evaluation.

ncRNAs: Databases and Prediction Programs

Noncoding RNAs (ncRNAs) are functional RNA molecules that are transcribed but not translated into proteins. Noncoding RNAs are highly abundant and functionally important RNAs such as transfer RNA (tRNAs) and ribosomal RNA (rRNAs), as well as snoRNAs, microRNAs, siRNAs, snRNAs, exRNAs, piRNAs and long ncRNAs (Amaral et al. 2008; Jacquier 2009; Ponting et al. 2009; Waters and Storz 2009; Liu and Camilli 2010). The precise amount of ncRNA molecules in any organism is unknown but is thought to be in the order of thousands, and many genome annotations list them together with protein-coding genes.

Many significant bioinformatics tools exist for protein-coding gene detection and analysis, and they are commonly used during genome annotations. In contrast, ncRNA gene detection is a much more complex problem. The main barrier to systematic de novo prediction of ncRNAs is that there are no common statistically significant features in their primary sequence (e.g., open reading frames or codon bias) which could be exploited by algorithms.

The most complete online database for ncRNAs is RFAM (Burge et al. 2013). RFAM 11.0 hosts over 2,208 ncRNA families and provides tools to perform similarity searches against the database. Users can also access precomputed full genome annotations. The limitations of the RFAM website are the small size-limit for online searches and the restriction of matches to existing RNA families, which excludes a large number of ncRNAs for which no alignment is available or submitted to RFAM. In addition, the information stored about fungal ncRNA is very sparse, and covers few species, primarily *S. cerevisae* and *Aspergillus fumigatus* (Kavanaugh and Dietrich 2009; Jochl et al. 2008).

The majority of ncRNA databases focus on one or more specific class of RNA molecule, like miRBase for microRNAs (Griffiths-Jones 2010); one class for a set of organisms such as the snoRNABase (Lestrade and Weber 2006); or one genome such as *Arabidopsis* with the ASRP (http://asrp.cgrb.oregonstate.edu). Some, however, aim to providing an exhaustive catalog of ncRNAs such as NONCODE

(Bu et al. 2012), fRNAdb (Kin et al. 2007) or the comprehensive annotations of eukaryotic long noncoding RNAs found in lncRNAdb (Amaral et al. 2011). Details of the most common ncRNA databases available online are provided in Table 15.2.

Besides the functionality provided by the RFAM database site, several additional computational methods exist to perform ncRNA prediction on genomic sequences. These can be classified into two types, namely, those that aim to search for ncRNA homologs using conserved sequence and structure characteristics, and those that aim to discover new ncRNA families. Methods of the former type use sequence/structure alignments in order to identify key conserved motifs involved in the molecule structure. Alignments are then processed and modeled in several ways, for instance, with position weight matrices in ERPIN (Lambert et al. 2004) or covariance models in INFERNAL (Nawrocki and Eddy 2013). These types of methods also tend to be focused on one type of ncRNA, such as tRNAScan-SE (Schattner et al. 2005) and RNAmmer (Lagesen et al. 2007). Programs such as RNAMOTIF (Macke et al. 2001), CRT (Bland et al. 2007), and DARN! (Gaspin et al. 2008) provide users with a programming language allowing the description of any ncRNA structure.

A ready-to-use bioinformatics prediction protocol for various kinds of known ncRNAs is described at the Ensemble Fungi web site (http://fungi.ensembl.org/info/genome/genebuild/ncrna.html).

While generally ncRNA prediction is difficult and computationally expensive, a subclass of ncRNAs can—with the help of comparative genomics—be predicted with fair accuracy. Structural ncRNAs have a defined and evolutionary conserved secondary structure which is of functional importance. RNAz is a program for predicting structurally conserved and thermodynamically stable RNA secondary structures in multiple sequence alignments (Washietl et al. 2005). It can be used in genome-wide screens to detect functional RNA structures, as found in ncRNAs and cis-acting regulatory elements of mRNAs. RNAz provides a ready-to-use pipeline to perform large genome scale scanning for ncRNAs. Comprehensive RNAz screens have been done for several genomes including *S. cerevisae* (Steigele et al. 2007). The most important step during an RNAz analysis is choosing a well-defined set of multiple alignments. There are a variety of programs available to produce genome-wide alignments, but for some organisms complete comparative multispecies alignments are available online. For *S. cerevisae*, the UCSC genome browser provides Multiz alignments up to seven *Saccharomyces* species (http://www.genome.ucsc.edu/). Similar data sets are available for higher eukaryotes.

When screening large genomic regions for ncRNAs the greatest source of error relates to specificity. There is a constant background signal of false positives, and the more sequences included in the analysis, the more false positives. Therefore, it is recommended to focus the analysis only on specific regions of interest (e.g., intergenic regions). Once selected, the next step consists of preprocessing a set of raw alignments. The RNAz package provides all the additional tools requires to slice the alignments in overlapping windows, checking for unrelated alignments,

Table 15.2 Summary of commonncRNA online databases

Name	Description	Species specific	ncRNA specific	Last update or release
Rfam	A collection of RNA families, each represented by multiple sequence alignments, consensus secondary structures, and covariance models	No	no	August 2012
miRBase	A searchable database of published miRNA sequences and annotation.	No	miRNA	June 2013
snoRNA Base	A comprehensive database of human H/ACA and C/D box snoRNAs	*H. sapiens*	snoRNA	July 2007
ASRP	The *Arabidopsis thaliana* small RNA project	A. Thaliana	smallRNA	June 2013
Non code	A database of all types of noncoding RNAs (except tRNAs and rRNAs)	No	All except tRNA and rRNA	October 2013
fRNAdb	A database of comprehensive ncRNA sequences acquired from other sequence databases	No	no	–
lncRNAdb	A database of long noncoding RNA which have been experimentally characterized with a biological function	No	lncRNA	January 2011
ChIPBase	An integrated resource and platform for decoding transcription factor binding maps, expression profiles and transcriptional regulation of noncoding RNAs and protein-coding genes from ChIP-Seq data	No	Chip-Seq data	November 2012
RNAdb	A database of mammalian noncoding RNAs	Mammals	no	Retired in June 2012
Plant snoRNA database	A database of plant Small nucleolar RNAs	Plants	snoRNA	–
Yeast snoRNA database	A comprehensive database of S. cerevisiae H/ACA and C/D box snoRNAs	Yeast	snoRNA	April 2010
tmRDB	A database which provides aligned, annotated and phylogenetically-ordered sequences related to structure and function of Transfer-messenger RNA	No	tmRNP	–

and discarding duplicate sequences or sequences with high GC content. Once the alignments have been prepared, RNAz can be run to score them, cluster them into loci and filter the output. The results can be viewed as a webpage, BED file, or GFF annotation.

The Vienna RNA Package has been a widely used compilation of RNA secondary structure-related tools and algorithms for nearly two decades (Hofacker 2003), it consists of a C code library and several standalone programs for the prediction and comparison of RNA secondary structures. RNA secondary structure prediction through energy minimization is the most used function in the package. Several other programs are included in the package, for instance, tools to evaluate energy of RNA secondary structures, predict minimum energy secondary structures and pair probabilities, predict the consensus structure of several aligned sequences, and many others.

References

Ahmed F, Kumar M, Raghava GPS (2009) Prediction of polyadenylation signals in human DNA sequences using nucleotide frequencies. In Silico Biol 9:135–148

Amaral PP, Clark MB, Gascoigne DK, Dinger ME, Mattick JS (2011) lncRNAdb: a reference database for long noncoding RNAs. Nucleic Acids Res 39:D146–D151

Amaral PP, Dinger ME, Mercer TR, Mattick JS (2008) The eukaryotic genome as an RNA machine. Sci (NY) 319:1787–1789

Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11(10):R106. doi:10.1186/gb-2010-11-10-r106

Arvas M, Pakula T, Smit B, Rautio J, Koivistoinen H, Jouhten P, Lindfors E, Wiebe M, Penttila M, Saloheimo M (2011) Correlation of gene expression and protein production rate—a system wide study. BMC Genom 12:616. doi:10.1186/1471-2164-12-616

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 37:W202–W208

Beck AH, Weng Z, Witten DM, Zhu S, Foley JW, Lacroute P, Smith CL, Tibshirani R, van de Rijn M, Sidow A, West RB (2010) 3rd end. Sequencing for expression quantification (3SEQ) from archival tumor samples. PLoS ONE 5:e8768

Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC, Hugenholtz P (2007) CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. BMC Bioinformatics 8:209. doi:10.1186/1471-2105-8-209

Brockman JM, Singh P, Liu D, Quinlan S, Salisbury J, Graber JH (2005) PACdb: PolyA cleavage site and 3′-UTR database. Bioinformatics, vol 21. Oxford, England, p 3691–3693

Bu D, Yu K, Sun S, Xie C, Skogerb\o G, Miao R, Xiao H, Liao Q, Luo H, Zhao G, Zhao H, Liu Z, Liu C, Chen R, Zhao Y (2012) NONCODE v3.0: integrative annotation of long noncoding RNAs. Nucleic Acids Res 40:D210–215

Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP, Eddy SR, Gardner PP, Bateman A (2013) Rfam 11.0: 10 years of RNA families. Nucleic Acids Res 41:D226–232

Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJ (2009) Biopython: freely available python tools for computational molecular biology and bioinformatics. Bioinformatics 25(11):1422–1423. doi:10.1093/bioinformatics/btp163

Cock PJ, Fields CJ, Goto N, Heuer ML, Rice PM (2010) The sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res 38(6):1767–1771. doi:10.1093/nar/gkp1137

Chen LL, Carmichael GG (2010) Long noncoding RNAs in mammalian cells: what, where, and why? Wiley Interdiscip Rev RNA 1:2–21

Cheng Y, Miura RM, Tian B (2006) Prediction of mRNA polyadenylation sites by support vector machine. Bioinformatics (Oxford, England) 22:2320–2325

David M, Dzamba M, Lister D, Ilie L, Brudno M (2011) SHRiMP2: sensitive yet practical SHort Read Mapping. Bioinformatics 27(7):1011–1012. doi:10.1093/bioinformatics/btr046

Derti A, Garrett-Engele P, Macisaac KD, Stevens RC, Sriram S, Chen R, Rohl CA, Johnson JM, Babak T (2012) A quantitative atlas of polyadenylation in five mammals. Genome Res 22(6):1173–1183. doi:10.1101/gr.132563.111

Di Giammartino DC, Nishida K, Manley JL (2011) Mechanisms and consequences of alternative polyadenylation. Mol Cell 43(6):853–866. doi:10.1016/j.molcel.2011.08.017

Elahi E, Ronaghi M (2004) Pyrosequencing: a tool for DNA sequencing analysis. Methods Mol Biol 255:211–219. doi:10.1385/1-59259-752-1:211

Flavell SW, Kim T-K, Gray JM, Harmin DA, Hemberg M, Hong EJ, Markenscoff-Papadimitriou E, Bear DM, Greenberg ME (2008) Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. Neuron 60:1022–1038

Fox-Walsh K, Davis-Turak J, Zhou Y, Li H, Fu X-D (2011) A multiplex RNA-seq strategy to profile poly(A+) RNA: application to analysis of transcription response and 3rd edn formation. Genomics 98:266–271

Gaspin C, Schiex T, Zytnicki M (2008) DARN! A weighted constraint solver for RNA motif localization. 10.1007/s10601-007-9033-9. 13

Griffiths-Jones S (2010) miRBase: microRNA sequences and annotation. Current protocols in bioinformatics / editoral board. Andreas D Baxevanis et al (eds) Chapter 12: Unit 12.19.11-10

Grillo G, Turi A, Licciulli F, Mignone F, Liuni S, Banfi S, Gennarino VA, Horner DS, Pavesi G, Picardi E, Pesole G (2010) UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. Nucleic Acids Res 38:D75–D80

Hofacker I (2003) Vienna RNA secondary structure server. Nucleic Acids Res 31:3429–3431

Hoque M, Ji Z, Zheng D, Luo W, Li W, You B, Park JY, Yehia G, Tian B (2013) Analysis of alternative cleavage and polyadenylation by 3′ region extraction and deep sequencing. Nat Methods 10:133–139

Jacquier A (2009) The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. Nat Rev Genet 10:833–844

Jan CH, Friedman RC, Ruby JG, Bartel DP (2011) Formation, regulation and evolution of Caenorhabditis elegans 3′UTRs. Nature 469:97–101

Jochl C, Rederstorff M, Hertel J, Stadler PF, Hofacker IL, Schrettl M, Haas H, Huttenhofer A (2008) Small ncRNA transcriptome analysis from Aspergillus fumigatus suggests a novel mechanism for regulation of protein synthesis. Nucleic Acids Res 36(8):2677–2689. doi:10.1093/nar/gkn123

Kalkatawi M, Rangkuti F, Schramm M, Jankovic BR, Kamau A, Chowdhary R, Archer JAC, Bajic VB (2012) Dragon PolyA spotter: predictor of poly(A) motifs within human genomic DNA sequences. Bioinformatics (Oxford, England) 28:127–129

Kavanaugh LA, Dietrich FS (2009) Non-coding RNA prediction and verification in Saccharomyces cerevisiae. PLoS Genet 5:e1000321

Kin T, Yamada K, Terai G, Okida H, Yoshinari Y, Ono Y, Kojima A, Kimura Y, Komori T, Asai K (2007) fRNAdb: a platform for mining/annotating functional RNA candidates from non-coding RNA sequences. Nucleic Acids Res 35:D145–D148

Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner DJ (2009) Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. Nat Methods 6(4):291–295. doi:10.1038/nmeth.1311

Lagesen K, Hallin P, R\o dland EA, Staerfeldt H-H, Rognes T, Ussery DW (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35:3100–3108

Lambert AE, Fontaine J-F, Legendre M, Leclerc F, Permal E, Major FC, Putzer H, Delfour O, Michot B, Gautheret D (2004) The ERPIN server: an interface to profile-based RNA motif identification. Nucleic Acids Res. 32:W160–W165

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with bowtie 2. Nat Methods 9(4):357–359. doi:10.1038/nmeth.1923

Lee JY, Yeh I, Park JY, Tian B (2007) PolyA_DB 2: mRNA polyadenylation sites in vertebrate genes. Nucleic Acids Res 35:D165–D168

Lestrade L, Weber MJ (2006) snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. Nucleic Acids Res 34:D158–D162

Li H, Durbin R (2010) Fast and accurate long-read alignment with burrows-wheeler transform. Bioinformatics 26(5):589–595. doi:10.1093/bioinformatics/btp698

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S (2009) The sequence alignment/map format and SAMtools. bioinformatics 25(16):2078–2079. doi:10.1093/bioinformatics/btp352

Liu JM, Camilli A (2010) A broadening world of bacterial small RNAs. Curr Opin Microbiol 13:18–23

Lunter G, Goodson M (2011) Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res 21(6):936–939. doi:10.1101/gr.111120.110

Lutz CS, Moreira A (2010) Alternative mRNA polyadenylation in eukaryotes: an effective regulator of gene expression. Wiley Interdisc Rev RNA 2:22–31

Macke TJ, Ecker DJ, Gutell RR, Gautheret D, Case DA, Sampath R (2001) RNAMotif, an RNA secondary structure definition and search algorithm. Nucleic Acids Res 29:4724–4735

Mardis ER (2008) Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet 9:387–402. doi:10.1146/annurev.genom.9.081307.164359

M-Je Schmidt, Norbury CJ (2010) Polyadenylation and beyond: emerging roles for noncanonical poly(A) polymerases. Wiley interdisc Rev RNA 1:142–151

McKernan KJ, Peckham HE, Costa GL, McLaughlin SF, Fu Y, Tsung EF, Clouser CR, Duncan C, Ichikawa JK, Lee CC, Zhang Z, Ranade SS, Dimalanta ET, Hyland FC, Sokolsky TD, Zhang L, Sheridan A, Fu H, Hendrickson CL, Li B, Kotler L, Stuart JR, Malek JA, Manning JM, Antipova AA, Perez DS, Moore MP, Hayashibara KC, Lyons MR, Beaudoin RE, Coleman BE, Laptewicz MW, Sannicandro AE, Rhodes MD, Gottimukkala RK, Yang S, Bafna V, Bashir A, MacBride A, Alkan C, Kidd JM, Eichler EE, Reese MG, De La Vega FM, Blanchard AP (2009) Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. Genome Res 19(9):1527–1541. doi:10.1101/gr.091868.109

Mischo HE (1829) Proudfoot NJ (2013) Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast. Biochim Biophys Acta 1:174–185. doi:10.1016/j.bbagrm.2012.10.003

Morris AR, Bos A, Diosdado B, Rooijers K, Elkon R, Bolijn AS, Carvalho B, Meijer GA, Agami R (2012) Alternative cleavage and polyadenylation during colorectal cancer development. Clin Cancer Res Official J Am Assoc Cancer Res 18(19):5256–5266. doi:10.1158/1078-0432.CCR-12-0543

Nam DK, Lee S, Zhou G, Cao X, Wang C, Clark T, Chen J, Rowley JD, Wang SM (2002) Oligo(dT) primer generates a high frequency of truncated cDNAs through internal poly(A) priming during reverse transcription. Proc Natl Acad Sci USA 99:6152–6156

Nawrocki EP, Eddy SR (2013) Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics Oxford, England p 1–3

Ozsolak F, Kapranov P, Foissac S, Kim SW, Fishilevich E, Monaghan AP, John B, Milos PM (2010) Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. Cell 143:1018–1029

Pelechano V, Wilkening S, Järvelin AI, Tekkedil MM, Steinmetz LM (2012) Genome-wide polyadenylation site mapping. Meth Enzymol 513:271–296. doi:10.1016/B978-0-12-391938-0.00012-4

Perez-Canadillas JM (2006) Grabbing the message: structural basis of mRNA 3′UTR recognition by Hrp1. EMBO J 25(13):3167–3178. doi:10.1038/sj.emboj.7601190

Ponting CP, Oliver PL, Reik W (2009) Evolution and functions of long noncoding RNAs. Cell 136:629–641

Reimers M, Carey VJ (2006) Bioconductor: an open source framework for bioinformatics and computational biology. Methods Enzymol 411:119–134. doi:10.1016/S0076-6879(06)11008-3

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26(1):139–140. doi:10.1093/bioinformatics/btp616

Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB (2008) Proliferating cells express mRNAs with shortened 3′ untranslated regions and fewer micro RNA target sites. Sci NY 320:1643–1647

Schattner P, Brooks AN, Lowe TM (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res 33:W686–W689

Shen Y, Liu Y, Liu L, Liang C, Li QQ (2008) Unique features of nuclear mRNA poly(A) signals and alternative polyadenylation in Chlamydomonas reinhardtii. Genetics 179(1):167–176. doi:10.1534/genetics.108.088971

Shendure J, Ji H (2008) Next-generation DNA sequencing. Nat Biotechnol 26(10):1135–1145. doi:10.1038/nbt1486

Shepard PJ, Choi E-A, Lu J, Flanagan LA, Hertel KJ, Shi Y (2011) Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. RNA NY 17:761–772

Singh P, Alley TL, Wright SM, Kamdar S, Schott W, Wilpan RY, Mills KD, Graber JH (2009) Global changes in processing of mRNA 3′ untranslated regions characterize clinically distinct cancer subtypes. Cancer Res 69:9422–9430

Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigian C, Fuellen G, Gilbert JG, Korf I, Lapp H, Lehvaslaiho H, Matsalla C, Mungall CJ, Osborne BI, Pocock MR, Schattner P, Senger M, Stein LD, Stupka E, Wilkinson MD, Birney E (2002) The bioperl toolkit: perl modules for the life sciences. Genome Res 12(10):1611–1618. doi:10.1101/gr.361602

Steigele S, Huber W, Stocsits C, Stadler PF, Nieselt K (2007) Comparative analysis of structured RNAs in S. cerevisiae indicates a multitude of different functions. BMC Biol 5:25. doi:10.1186/1741-7007-5-25

Thomas-Chollier M, Defrance M, Medina-Rivera A, Sand O, Herrmann C, Thieffry D, van Helden J (2011) RSAT 2011: regulatory sequence analysis tools. Nucleic Acids Res 39:W86–W91

Tian B, Hu J, Zhang H, Lutz CS (2005) A large-scale analysis of mRNA polyadenylation of human and mouse genes. Nucleic Acids Res 33:201–212

Trapnell C, Pachter L, Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25(9):1105–1111. doi:10.1093/bioinformatics/btp120

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and cufflinks. Nat Protoc 7(3):562–578. doi:10.1038/nprot.2012.016

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB (2008) Alternative isoform regulation in human tissue transcriptomes. Nature 456:470–476

Wang K, Singh D, Zeng Z, Coleman SJ, Huang Y, Savich GL, He X, Mieczkowski P, Grimm SA, Perou CM, MacLeod JN, Chiang DY, Prins JF, Liu J (2010) MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. Nucleic Acids Res 38(18):e178. doi:10.1093/nar/gkq622

Wang KC, Chang HY (2011) Molecular mechanisms of long noncoding RNAs. Mol Cell 43(6):904–914. doi:10.1016/j.molcel.2011.08.018

Washietl S, Hofacker IL, Stadler PF (2005) Fast and reliable prediction of noncoding RNAs. Proc Natl Acad Sci USA 15:2454–2459. doi:10.1073/pnas.0409169102

Waters LS, Storz G (2009) Regulatory RNAs in bacteria. Cell 136:615–628

Wilkening S, Pelechano V, J\"arvelin AI, Tekkedil MM, Anders S, Benes V, Steinmetz LM (2013) An efficient method for genome-wide polyadenylation site mapping and RNA quantification. Nucleic Acids Res 41:e65

Wilson JT, deRiel JK, Forget BG, Marotta CA, Weissman SM (1977) Nucleotide sequence of 3′ untranslated portion of human alpha globin mRNA. Nucleic Acids Res 4(7):2353–2368

Wu TD, Nacu S (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 26(7):873–881. doi:10.1093/bioinformatics/btq057

Yoon OK, Brem RB (2010) Noncanonical transcript forms in yeast and their regulation during environmental stress. RNA NY 16:1256–1267

Zhang H, Lee JY, Tian B (2005) Biased alternative polyadenylation in human tissues. Genome Biol 6:R100

Index

A. Sesma and T. von der Haar (eds.), *Fungal RNA Biology*,
DOI: 10.1007/978-3-319-05687-6,