Topics in Current Chemistry 330

Jonathan B. Chaires
David Graves *Editors*

Quadruplex Nucleic Acids

Springer

330
Topics in Current Chemistry

For further volumes:
http://www.springer.com/series/128

Aims and Scope

The series Topics in Current Chemistry presents critical reviews of the present and future trends in modern chemical research. The scope of coverage includes all areas of chemical science including the interfaces with related disciplines such as biology, medicine and materials science.

The goal of each thematic volume is to give the non-specialist reader, whether at the university or in industry, a comprehensive overview of an area where new insights are emerging that are of interest to larger scientific audience.

Thus each review within the volume critically surveys one aspect of that topic and places it within the context of the volume as a whole. The most significant developments of the last 5 to 10 years should be presented. A description of the laboratory procedures involved is often useful to the reader. The coverage should not be exhaustive in data, but should rather be conceptual, concentrating on the methodological thinking that will allow the non-specialist reader to understand the information presented.

Discussion of possible future research directions in the area is welcome.

Review articles for the individual volumes are invited by the volume editors.

Readership: research chemists at universities or in industry, graduate students.

Jonathan B. Chaires • David Graves
Editors

Quadruplex Nucleic Acids

With contributions by

R. Buscaglia · J.B. Chaires · W.L. Dean · A. De Cian · V. Dötsch · S. Foldynová-Trantírková · T. Fujimoto · C. Giancola · V. Gokhale · A. Granzhan · J. Gros · F. Hamon · R. Hänsel · L.H. Hurley · C.E. Kaiser · E. Largy · H.T. Le · J.-L. Mergny · D. Miyoshi · R. Moriyama · B. Pagano · L. Petraccone · A. Randazzo · G.P. Spada · N. Sugimoto · M.-P. Teulade-Fichou · P.L.T. Tran · L. Trantírek · J.O. Trent · D. Verga · M. Webba da Silva · D. Yang

Springer

Editors
Jonathan B. Chaires
James Graham Brown Cancer Center
University of Louisville
Louisville, KY
USA

David Graves
Department of Chemistry
University of Alamaba at Birmingham
Birmingham, Alaska
USA

ISSN 0340-1022 ISSN 1436-5049 (electronic)
ISBN 978-3-642-34742-9 ISBN 978-3-642-34743-6 (eBook)
DOI 10.1007/978-3-642-34743-6
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012954486

Printed on acid-free paper

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

This volume is dedicated to Donald M. Miller, MD, PhD, in appreciation of his leadership, encouragement, and support

Preface

G-quadruplexes have advanced from being a biophysical oddity to being at the forefront of chemical biology research. Quadruplexes form as the result of the unique ability of guanine to form an array of specific hydrogen bonds to self-assemble into a square planar quartet structure. These G-quartets can then further assemble by stacking to form a variety of stable G-quadruplex structures that feature diverse molecularity, topologies, and strand-segment polarities depending on the exact nucleic acid sequences involved. Both DNA and RNA can form quadruplex structures. While the chemical and physical properties of G-quadruplexes are fascinating on their own, seminal work in the 1980s from the Blackburn, Cech, Klug, and Gilbert laboratories suggested that quadruplexes might in addition play important functional roles in biology, particularly in telomere replication and regulation and in recombination. Subsequent bioinformatics studies from a number of laboratories have shown that sequences that might form quadruplex structures are highly conserved and nonrandomly positioned throughout the human genome and the genomes of a number of other eukaryotes, pointing again to a probable functional role for quadruplexes. Quadruplex research has undergone explosive growth since, such that there are now over 400 new publications on quadruplexes per year, and over 12,000 citations per year to the quadruplex literature. Interest in quadruplexes spans diverse disciplines, including chemistry, biophysics, molecular and cellular biology, drug discovery, and nanotechnology.

This volume in the *Topics in Current Chemistry* series provides several snapshots of cutting-edge research in the vibrant quadruplex field. A completely comprehensive and balanced compilation that would cover all aspects of contemporary quadruplex research would require more than a single volume. The selection of topics presented here stemmed from a personal view of key developments in the field, along with the recruitment of authors willing to pause to critically review the progress in their particular areas of interest.

Diverse areas of active research are represented in the volume. G-quadruplexes have emerged as critical elements in the control of gene expression. The contribution from combined laboratories of Laurence Hurley and Danzhou Yang describes the role of quadruplexes in the control of the c-MYC oncogene, and discusses the design of small molecules that might mediate expression of that gene by stabilizing the quadruplex. Higher-order quadruplex structures that might form in telomeric DNA are described by Luigi Petraccone. The structures he describes are beyond

what contemporary NMR and x-ray crystallographic methods can determine with high-resolution, and are instead inferred from a powerful integrated approach that combines molecular dynamics simulations coupled to rigorous experimental validation of predicted properties based on computed models. The chemical environment within cells surely differs from the dilute electrolyte solutions typically used in vitro to study quadruplex structures. Two chapters describe novel approaches to study quadruplexes under conditions that more realistically mimic the biological environment. The chapter by Hänsel and coworkers describes an exciting in-cell NMR method using a cell extract that permits exploration of nucleic acids structures under the crowded condition within cells. The chapter by Miyoshi and colleagues presents another, more traditional, approach that uses added cosolutes to mimic the effects of macromolecular crowding and changes in water activity on quadruplex structures.

Technological advances allow deeper understanding of quadruplex structures. Three chapters describe some of these advances. The chapter by Randazzo and colleagues presents new insights into the origin of the circular dichroism spectra of quadruplex structures. These insights place the interpretation of circular dichroism spectra on a much more firm physical chemical foundation. The chapter from the Teulade-Fichou laboratory describes the development of a comprehensive set of novel small molecule fluorescent probes that can selectively recognize quadruplex structures. These provide important new analytical tools for the detection of quadruplexes both in vivo and in vitro. Finally, the utility of hydrodynamic methods and computations for the study of quadruplexes is critically evaluated in a contribution from the Trent laboratory. The power of analytical ultracentrifugation to provide reliable structural information is well illustrated in these studies.

The volume includes descriptions of two physical chemical approaches that provide mechanistic details of quadruplex reactions. Kinetic studies that unravel the mechanism of tetramolecular quadruplex assembly are presented by the team from the Mergny laboratory, studies that may be most relevant for the nanotechnology community. The energetics of ligand binding to quadruplexes as determined by isothermal titration calorimetry is nicely summarized by the contribution from the Giancola laboratory. These studies provide a thermodynamic foundation for drug development and discovery.

The chapters presented here capture the diversity of quadruplex research and rapid growth and development of the field. I thank the authors for their hard work and diligence in the preparation of their contributions. I also thank the publishers for their invitation to highlight quadruplex research in their excellent series of *Topics in Current Chemistry*, and to Karin Bartsch and Elizabeth Hawkins for their assistance in the preparation of this volume.

Louisville, KY J. B. Chaires
2012

Contents

Gaining Insights into the Small Molecule Targeting of the G-Quadruplex in the c-MYC Promoter Using NMR and an Allele-Specific Transcriptional Assay 1
Christine E. Kaiser, Vijay Gokhale, Danzhou Yang, and Laurence H. Hurley

Higher-Order Quadruplex Structures 23
Luigi Petraccone

Investigation of Quadruplex Structure Under Physiological Conditions Using In-Cell NMR 47
Robert Hänsel, Silvie Foldynová-Trantírková, Volker Dötsch, and Lukáš Trantírek

Circular Dichroism of Quadruplex Structures 67
Antonio Randazzo, Gian Piero Spada, and Mateus Webba da Silva

Molecular Crowding and Hydration Regulating of G-Quadruplex Formation 87
Daisuke Miyoshi, Takeshi Fujimoto, and Naoki Sugimoto

Visualizing the Quadruplex: From Fluorescent Ligands to Light-Up Probes 111
Eric Largy, Anton Granzhan, Florian Hamon, Daniela Verga, and Marie-Paule Teulade-Fichou

Calculation of Hydrodynamic Properties for G-Quadruplex Nucleic Acid Structures from *in silico* Bead Models 179
Huy T. Le, Robert Buscaglia, William L. Dean, Jonathan B. Chaires, and John O. Trent

Energetics of Ligand Binding to G-Quadruplexes 211
Concetta Giancola and Bruno Pagano

Tetramolecular Quadruplex Stability and Assembly 243
Phong Lan Thao Tran, Anne De Cian, Julien Gros, Rui Moriyama, and Jean-Louis Mergny

Index 275

Top Curr Chem (2013) 330: 1–22
DOI: 10.1007/128_2012_333

Published online: 3 July 2012

Gaining Insights into the Small Molecule Targeting of the G-Quadruplex in the c-MYC Promoter Using NMR and an Allele-Specific Transcriptional Assay

Christine E. Kaiser, Vijay Gokhale, Danzhou Yang, and Laurence H. Hurley

Abstract G-quadruplexes (four-stranded DNA secondary structures) are showing promise as new targets for anticancer therapies. Specifically, G-quadruplexes in the proximal promoter region of regulatory genes have the potential to act as silencer elements and thereby turn off transcription. Thus, compounds that are capable of binding to and stabilizing G-quadruplexes would be of great benefit. In this chapter we describe two recent studies from our labs. In the first case, we use NMR to elucidate the structure of a 2:1 complex between a small molecule and the G-quadruplex in the c-MYC promoter. In the second case, we use an allele-specific transcription assay to demonstrate that the effect of a G-quadruplex-interactive compound is mediated directly through the G-quadruplex. Finally, we use this information to propose models for the interaction of various small molecules with the c-MYC G-quadruplex.

Keywords G-quadruplex · c-MYC · gene regulation · NMR structure

C.E. Kaiser
College of Pharmacy, University of Arizona, 1703 E. Mabel Street, Tucson, AZ 85721, USA

V. Gokhale, D. Yang (✉), and L.H. Hurley (✉)
College of Pharmacy, University of Arizona, 1703 E. Mabel Street, Tucson, AZ 85721, USA

Arizona Cancer Center, 1515 N. Campbell Avenue, Tucson, AZ 85724, USA

BIO5 Institute, 1657 E. Helen Street, Tucson, AZ 85721, USA
e-mail: yangd@pharmacy.arizona.edu; hurley@pharmacy.arizona.edu

Contents

1 Introduction 2
2 NMR Determination of the Quindoline-i:c-MYC G-Quadruplex Complex 5
2.1 Solution Structure of the 2:1 Quindoline-i:c-MYC G-Quadruplex Complex 7
2.2 Different Binding Interactions Between the 3′-End and 5′-End Complexes 8
2.3 Comparisons with Other Ligand:G-Quadruplex Complexes 9
2.4 Insights into the Structure-Based Design of G-Quadruplex-Interactive Compounds 9
3 The CA46 Allele-Specific Transcriptional Assay 10
3.1 Identification of More Specific c-MYC G-Quadruplex-Interactive Compounds 10
3.2 Validation of GQC-05 as a c-MYC G-Quadruplex-Binding Compound 12
3.3 Demonstration in a Cellular System That the Downstream Effects of the G-Quadruplex-Interactive Compounds Are Mediated Directly Through the c-MYC G-Quadruplex and Not Indirectly Through Other G-Quadruplexes or Other Cellular Targets 12
3.4 Demonstration That the Activating Transcription Factors Were Displaced from the NHE III_1 as a Consequence of Drug Stabilization of the c-MYC G-Quadruplex ... 14
4 Molecular Modeling of GQC-05 and Quarfloxin Based on the Quindoline-i:c-MYC G-Quadruplex Complex 14
4.1 Modeling Methods 16
4.2 Comparison of Binding Interactions 17
5 Conclusions 18
References 19

1 Introduction

The initial target for anticancer drugs was DNA and, today, drugs that target DNA remain the mainstay of most chemotherapy regimens [1]. However, interest in developing DNA-targeting drugs has waned due to a lack of selectivity for a particular sequence or region. Additionally, new and more specific molecular targets such as kinases and cell surface receptors have been identified for cancer. Recently, the discovery that telomeres and some guanine-rich (G-rich) promoter regions can form four-stranded DNA secondary structures termed G-quadruplexes [2–6] has ushered in an opportunity for a new phase of more selective DNA-targeted therapeutics.

The ability of G-rich telomeres to form higher order DNA structures was described in 1988 [4]. Telomeres occur at the end of eukaryotic chromosomes and contain large numbers of simple guanine-rich tandem repeats. The enzyme telomerase contributes to cell immortalization by catalyzing telomere extension and is over-expressed in a number of human cancers. In 1997 telomerase was shown to be inhibited by ligands that interact with G-quadruplexes in the 3′-end of human telomeric DNA [7]. Since then, bioinformatics studies have been performed to investigate the incidence of putative G-quadruplex-forming sequences (GPQSs) throughout the genome [3, 8–10]. These studies revealed GPQSs to be enriched in the proximal promoter regions of regulatory genes, especially of proto-oncogenes, in an evolutionarily conserved manner [2, 3]. Figure 1 shows the frequency (or probability) of each nucleotide upstream or downstream of the transcriptional start

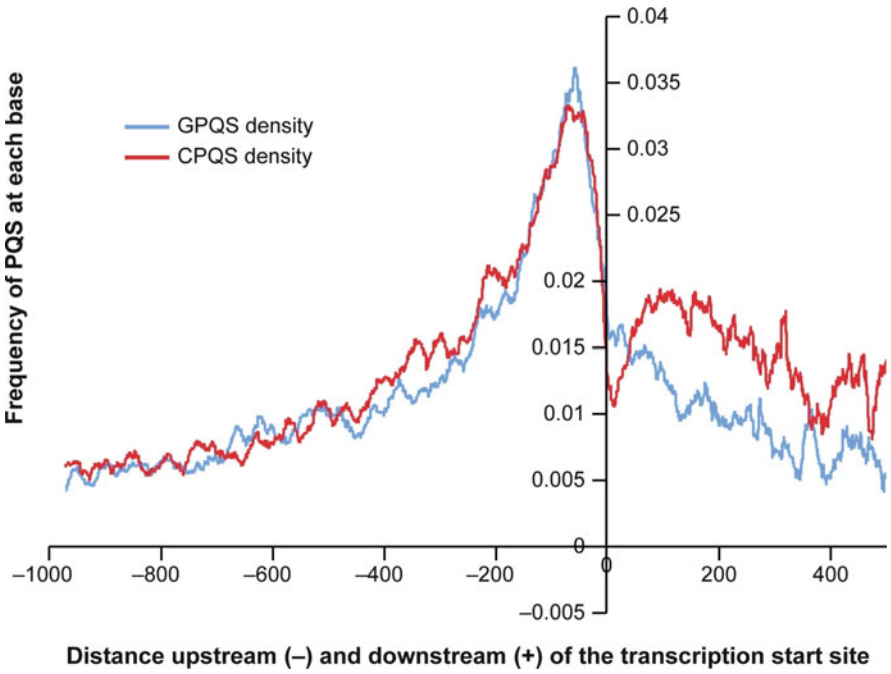


Fig. 1 Frequency plot showing the probability of each nucleotide upstream (−) or downstream (+) of the transcriptional start site being part of a putative G-quadruplex-forming sequence (GPQS), based on the Quadparser algorithm [9]. The data have been averaged over all human protein coding genes in the genome. The *blue plot* represents the presence of a G-quadruplex motif (GPQS) while the *red plot* represents the C-rich complement of a G-quadruplex motif (CPQS). Figure is modified, with permission, from [11], © (2008) Oxford University Press

site being part of a GPQS [10, 11]. This frequency peaks ~50 nucleotides upstream of the transcriptional start site, a region known for transcription regulation [10–12]. This clustering of GPQSs in known gene regulatory regions suggests that they may play a role in transcriptional control.

DNA supercoiling, a major source of superhelical stress in cells, is also known to play an important role in transcription [13]. As the transcriptional machinery translocates down a gene, it generates positive supercoiling downstream and negative supercoiling upstream. This negative supercoiling results in local unwinding of DNA and facilitates the opening of the guanine-rich/cytosine-rich (GC-rich) regions of the DNA. This process is thought to provide the equilibrating energy that facilitates the transition from double-stranded DNA to higher order non-B-DNA structures, allowing for the formation of G-quadruplexes on the G-rich strand and a complementary DNA secondary structure (i-motif) on the C-rich strand (Fig. 2) [15].

G-quadruplexes exhibit a diverse range of folding patterns: they are classified primarily by their tetrad directionality, loop length, and constitution [15]. These complex structures are highly polymorphic and capable of exhibiting

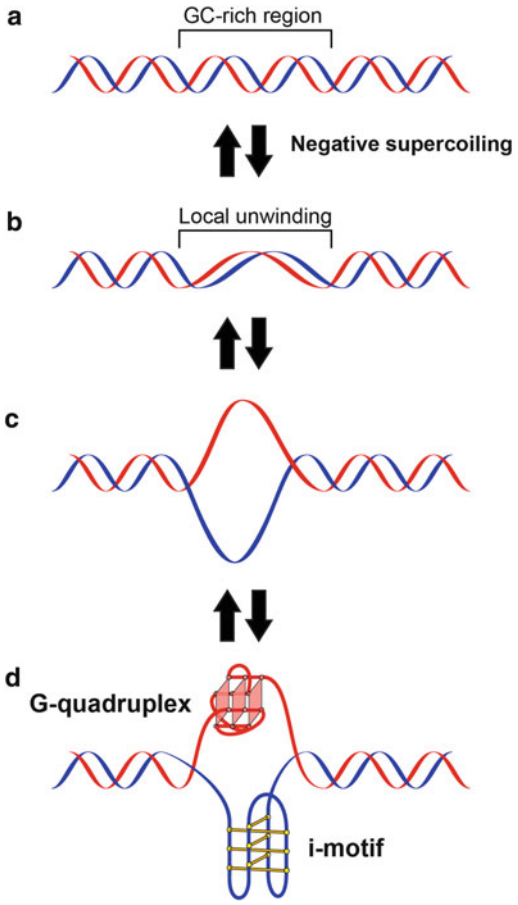


Fig. 2 Proposed equilibrating forms of DNA produced under negative supercoiling induced by transcription: (**a**) duplex, (**b**) locally unwound, (**c**) single stranded, (**d**) G-quadruplex/i-motif structure. Figure is modified, with permission, from [14], © (2009) American Chemical Society

parallel, antiparallel, or mixed topologies. The greatest sources of variability in G-quadruplexes are the lengths of their loops, which commonly range from one to nine base pairs, and their base composition, which has no constraints. On occasion, the loops themselves have been noted to form alternate DNA structures. For example, the G-quadruplex in the hTERT promoter contains a 26-base-pair loop observed to form a hairpin structure [15]. We have recently demonstrated that this tertiary DNA structure shows cooperativity in folding and unfolding such that a higher order interaction takes place between the hairpin and the G-quadruplex. This is the first example of cooperative refolding in DNA, a process previously documented only in proteins and complex RNA structures [16]. The unique globular shape and folded structures of G-quadruplexes and their potential ability to regulate the transcription of a host of oncogenes make them an attractive drug target.

Perhaps the most extensively studied and well-characterized promoter G-quadruplex is within the promoter of the c-MYC proto-oncogene, which plays a

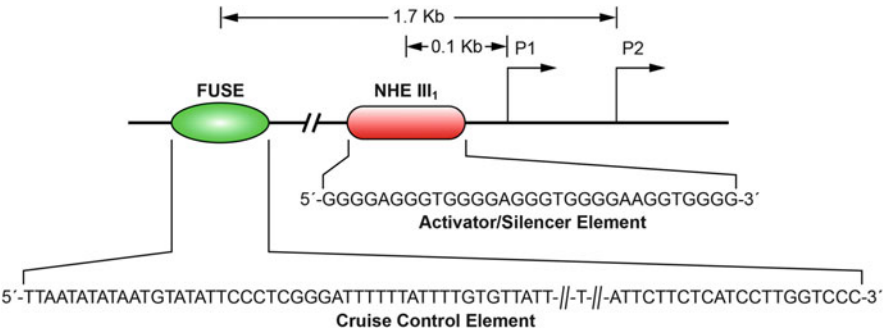


Fig. 3 The position and sequence of the FUSE element and NHE III_1 relative to the P1 and P2 promoters of c-MYC. The sequence shown for the FUSE element is only a partial representation. Figure is modified, with permission, from [18], © (2009) Macmillan Publishers Limited

role in normal cell proliferation and differentiation [17]. Transcriptional control of the c-MYC proto-oncogene has been investigated and described in numerous studies [2, 18, 19]. From these, two key regulatory regions have been identified: the far upstream element (FUSE) and the proximal nuclease hypersensitive element III_1 (NHE III_1) (Fig. 3). Using a mechanosensor mechanism fueled by transcriptionally induced negative supercoiling, the FUSE has been shown to function as a cruise control element [13]. The GC-rich NHE III_1, upstream of the core promoter region, is responsible for 80–90% of c-MYC transcription [19]. It consists of five repeats of the sequence TGGGGA(G/A)G(G/A) and serves as an activation/silencing region for c-MYC transcription [20]. Under negative supercoiling conditions, this region has been shown to form a G-quadruplex, which acts as a silencer element [14, 17, 21, 22].

c-MYC is aberrantly expressed, most often through an increase in transcription, in an estimated 80% of all human malignancies [17]. Thus, there is a considerable therapeutic value in the targeted down-regulation of c-MYC. In fact, Soucek and co-workers showed that systemic c-MYC inhibition in Ras-induced lung adenocarcinoma mouse models leads to regression of lung tumors [23]. Furthermore, while the systemic inhibition of c-MYC showed an effect on tissue regeneration, the effects were well tolerated and reversible [23]. With this in mind, we used NMR modeling and biological characterization to target the G-quadruplex in the c-MYC promoter in an attempt to find compounds that could down-regulate the expression of c-MYC at the transcriptional level.

2 NMR Determination of the Quindoline-i:c-MYC G-Quadruplex Complex

We previously determined the K^+ solution structure of the major G-quadruplex formed in the c-MYC promoter. This c-MYC promoter is a parallel-stranded structure with two G_3-N-G_3 single-nucleotide double-chain-reversal loops, represented by

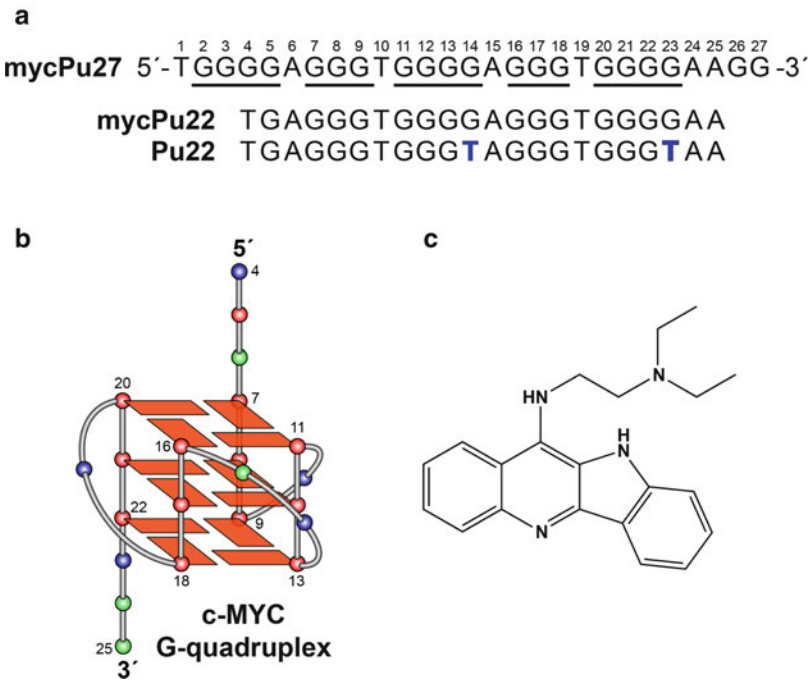


Fig. 4 (**a**) The promoter sequences of the NHE III_1 of the c-MYC gene and its modifications. mycPu27 is the wild-type 27-mer G-rich sequence of the c-MYC NHE III_1. mycPu22 is the wild-type 22-mer G-rich sequence of the c-MYC NHE III_1 that forms the major G-quadruplex in physiologically relevant K^+ solution. Pu22 is the modified mycPu22 sequence, with G-to-T substitutions at positions 14 and 23, that forms the predominant c-MYC promoter G-quadruplex in K^+ solution and whose structure was determined by NMR. (**b**) The folding topology of the c-MYC G-quadruplex adopted by Pu22. *Red box/red ball* = guanine; *green ball* = adenine, *blue ball* = thymine. (**c**) The quindoline-i molecule. Figure is modified, with permission, from [24], © (2011) American Chemical Society

the modified c-MYC promoter sequence, Pu22 (Fig. 4a, b) [25]. The G_3-N-G_3 motif has been shown to be a stable and prevalent structural motif in G-quadruplexes within a gene promoter [26]. Although the list of molecular structures reported for G-quadruplexes formed in gene promoters is growing, their structures with small molecules have been more difficult to obtain. We have very recently determined the NMR structure of a 2:1 complex of quindoline-i and the c-MYC G-quadruplex (PDB code: 2L7V) [24], which represents the first drug complex structure of a biologically relevant unimolecular promoter G-quadruplex.

The quindoline-i compound (Fig. 4c) is a derivative of the natural product cryptolepine [27, 28]. This compound has been shown to stabilize the G-quadruplex formed in the c-MYC promoter and subsequently inhibit the expression of c-MYC in the hepatocellular carcinoma cell line H2p G2 [29]. The structure showed an unexpected drug-induced reorientation of the flanking sequences at both ends of the DNA sequence. This reorientation, called an “induced intercalated triad pocket,”

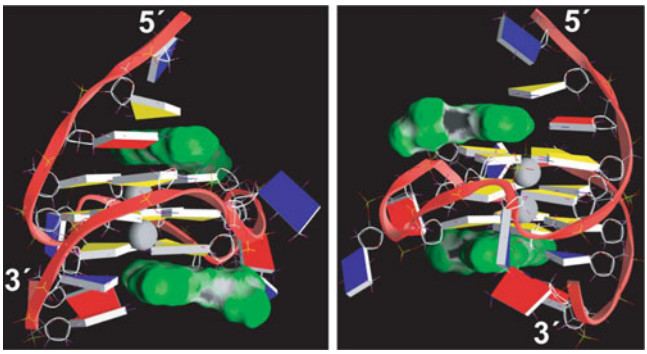


Fig. 5 A representative model of the NMR-refined 2:1 quindoline-i:c-MYC G-quadruplex complex structure from two different views, prepared using GRASP (guanine = *yellow*, adenine = *red*, thymine = *blue*). The quindoline-i molecules are shown as a space-filling model in *green*. The two potassium ions are shown as *gray balls*. Figure is modified, with permission, from [24], © (2011) American Chemical Society

acts as a mode of recognition and is analogous to the process used by riboswitches [30]. This mode of binding is distinct from previously proposed models wherein the planar G-quadruplex-interactive compounds such as telomestatin [31] and TMPyP4 [32] stack on the external guanine tetrad (G-tetrad). Additionally, the NMR structure of the quindoline-i:c-MYC G-quadruplex complex indicates that asymmetric compounds with a crescent shape, appropriate functional groups, and a small stacking moiety are more likely to bind in a defined manner to a unimolecular parallel-stranded G-quadruplex. This study describes the importance of the ligand shape as well as the two flanking bases of the G-quadruplex in determining drug-binding specificity and provides important insights for the structure-based rational design of molecules that interact with unimolecular parallel-stranded G-quadruplexes commonly found in promoter elements.

2.1 Solution Structure of the 2:1 Quindoline-i:c-MYC G-Quadruplex Complex

Two views showing representative structures of the 2:1 complex of quindoline-i and the c-MYC G-quadruplex are shown in Fig. 5. The quindoline-i-bound c-MYC G-quadruplex adopts the same folding pattern as that of the free DNA (see Fig. 4b) [25]. Instead of stacking on the external G-tetrads, both quindoline-i molecules bind to the c-MYC G-quadruplex in an "induced-fit" manner. Upon drug binding, both flanking sequences of the c-MYC G-quadruplex undergo an unexpected and large conformational change to assemble new capping structures. At both ends of the c-MYC G-quadruplex, the quindoline-i molecule stacks with just two of the four guanines of each external tetrad. Furthermore, the +1 thymine flanking base at the

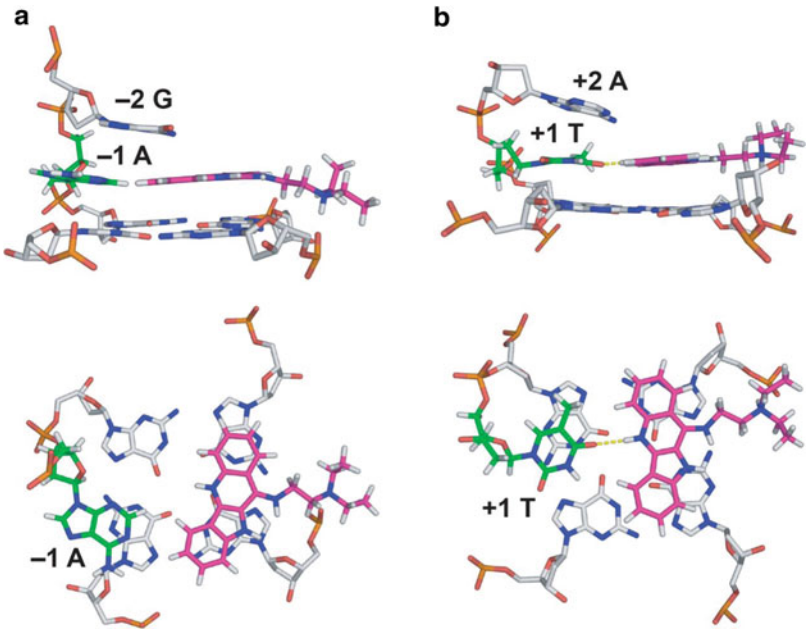


Fig. 6 Two different views of the drug-induced binding pockets at the 5′-end (**a**) and the 3′-end (**b**). The 3′- and 5′-end flanking bases are labeled. Figure is modified, with permission, from [24] © (2011) American Chemical Society

3′-end and the −1 adenine flanking base at the 5′-end are recruited to form a quasi-triad plane (Fig. 6). The +2 adenine and −2 guanine flanking residues wrap over the newly formed quasi-triad planes at each end (Fig. 6). This solution structure for the induced intercalated triad pocket shows a reorientation of the flanking sequence with the external tetrad and quindoline-i.

2.2 *Different Binding Interactions Between the 3′-End and 5′-End Complexes*

The 3′- and 5′-ends of the quindoline-i:c-MYC G-quadruplex complex have many features in common, including base stacking over two adjacent guanines and recruitment of either the −1 or +1 base that is aligned in the same plane as quindoline-i. However, the 3′- and 5′-ends also exhibit some important differences. For example, the 5′-face is more hydrophobic and more accessible for ligand stacking, whereas the 3′-face is more hydrophilic and less accessible for ligand stacking. Furthermore, the 5′-end complex is the more stable end. Its stability is highly dependent on the capping −2 guanine (mentioned above) since deletion of or mutation to a thymine dramatically reduces the stability of the complex (Fig. 6a).

The 3′-end complex is only stable under low ionic strength, emphasizing the importance of ionic rather than stacking interactions (Fig. 6b). The origin of the differences between the two ends is likely due to inherent structural features associated with the 3′- and 5′-faces as well as the flanking sequences.

2.3 Comparisons with Other Ligand:G-Quadruplex Complexes

Most of the known ligand:G-quadruplex complex structures have been derived from telomeric sequences that form bimolecular and tetramolecular species and determined by X-ray crystallography [33]. Prior to this work, TMPyP4, in complex with the unimolecular c-MYC promoter G-quadruplex, was the only known NMR-derived ligand:G-quadruplex structure for a promoter region [34]. However, the modified c-MYC promoter sequence used in that study contains a guanine-to-inosine substitution. This substitution induces a guanine-strand discontinuity at a guanine position shown to be critical for G-quadruplex formation in the wild-type mycPu27 sequence [35]. In this complex, TMPyP4 stacks over the 5′-end; however, the orientation of TMPyP4 was not resolved by NMR data, and a well-defined binding pocket was not observed. There are two dimeric telomeric G-quadruplexes in which the recruitment of a base into the plane of the ligand similar to that demonstrated by our work occurs. In these structures, a thymine residue is recruited into an in-line triad plane [36, 37]; however, because of the multimeric nature of both structures, they are less relevant to the unimolecular species described here.

2.4 Insights into the Structure-Based Design of G-Quadruplex-Interactive Compounds

The 2:1 quindoline-i:c-MYC G-quadruplex complex structure provides an important case study for the selective binding of ligands to unimolecular parallel G-quadruplexes in the promoter elements of various genes such as c-MYC, VEGF, Hif-1α, and c-KIT [38]. An important implication from our structure is that, unlike the symmetrical cyclic ligand typified by TMPyP4, asymmetric compounds that contain a smaller stacking moiety, such as quindoline-i, show more selective binding to a unimolecular parallel G-quadruplex.

The specific binding of the quindoline-i is determined by both the identity of the binding end (3′ or 5′) and the two flanking bases. In addition, the electrostatic interaction between the diethylamino group in the side chain of quindoline-i and the DNA phosphate backbone could help orient and stabilize the quindoline-i scaffold. In turn, this electrostatic interaction may pinpoint the potential location of substituents that can interact with the loops of the G-quadruplex. Small changes in the shape or electronic structure of the ligand, or in the identity of the flanking bases, may affect the precise positioning of the ligand in relation to the G-quadruplex.

On the basis of our work, we propose a two-step process for small molecule recognition of unimolecular parallel G-quadruplexes by small molecules similar to

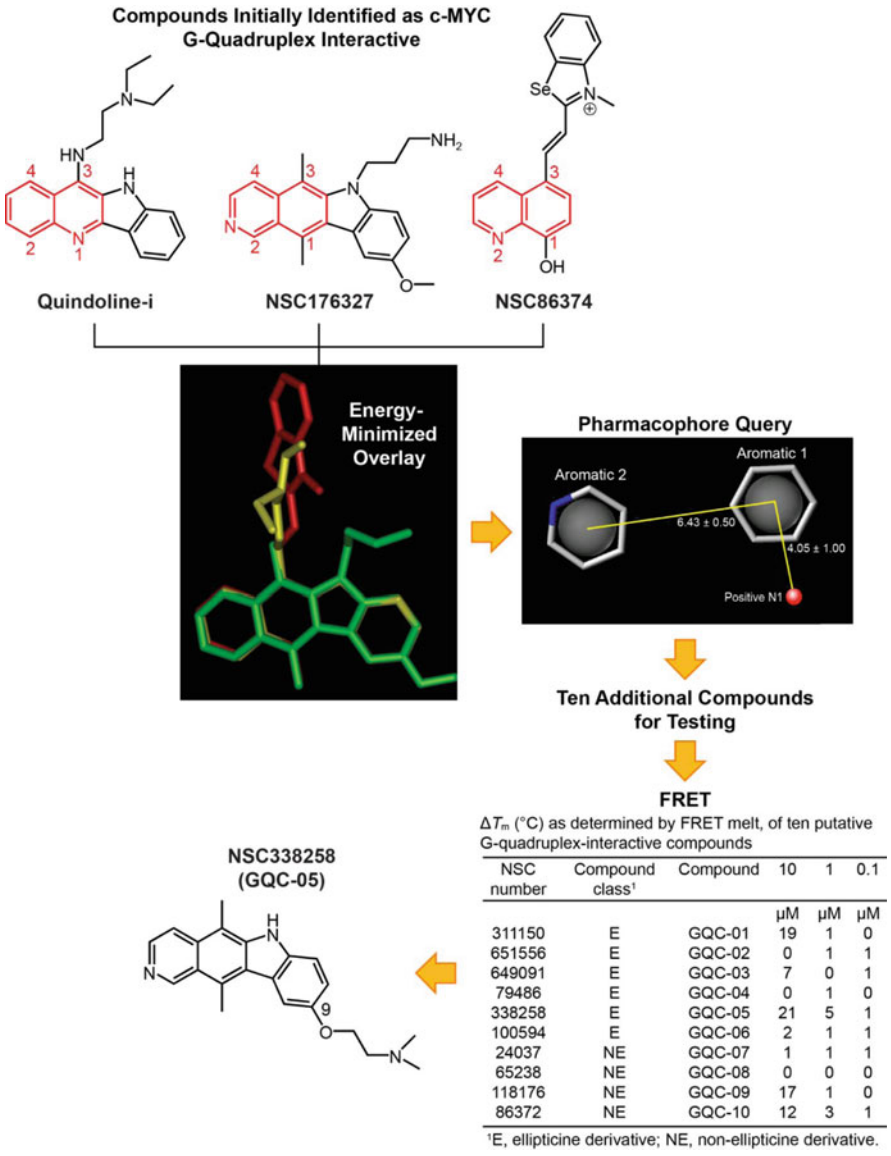
quindoline-i. First, the small molecule induces a large conformation change in the G-quadruplex flanking bases, which results in the formation of an intercalated triad pocket. Second, the substituents on the small molecule interact with the loops of the G-quadruplex Identification of this small molecule-induced rearrangement of the flanking sequence suggests that other ligands may produce their own unique binding pockets in promoter G-quadruplexes and thus provide unexpected opportunities for structure-based drug design.

3 The CA46 Allele-Specific Transcriptional Assay

While we were developing an NMR-derived structure of the quindoline-i:c-MYC G-quadruplex complex, we also sought to provide more convincing evidence for the biological role of G-quadruplexes and their drug targeting with small molecules similar to quindoline-i. Specifically, we were interested in finding evidence that the inhibition observed for ligand-mediated transcription of c-MYC was directly mediated through the G-quadruplex found in the NHE III_1 of the c-MYC promoter region and not through secondary off-target effects. Probably the most exacting test used prior to the recent publication in Journal of Biological Chemistry [17] was the comparison of effects of G-quadruplex-interactive compounds in two Burkitt's lymphoma cell lines, CA46 and Ramos. Only in the case of the Ramos cell line is the G-quadruplex retained in the translocated allele; therefore compounds that work directly through the c-MYC G-quadruplex should have a more profound effect on c-MYC transcription in this cell line than in CA46. Indeed, TMPyP4, but not TMPyP2 (its positional and less G-quadruplex-interactive isomer), appears to have this preferential effect. Likewise, quindoline-i and actinomycin also showed positive results in this assay [29, 39]. However, even in the translocated allele of the Ramos cell line, the possibility remains that the inhibitory effect on c-MYC was at least partially due to a secondary effect. To provide a more exacting system to determine whether G-quadruplex-interactive compounds work directly through the c-MYC G-quadruplex, we embarked on a project in which we attempted to (1) identify more specific c-MYC G-quadruplex-interactive compounds, (2) demonstrate in a cellular system that the downstream effects of G-quadruplex-interactive compounds are mediated directly through the c-MYC G-quadruplex and not indirectly through other G-quadruplexes or other cellular targets, and (3) demonstrate that the activating transcriptional factors were displaced from the NHE III_1 as a direct consequence of drug binding to the c-MYC G-quadruplex.

3.1 Identification of More Specific c-MYC G-Quadruplex-Interactive Compounds

The flowchart in Fig. 7 shows the procedure used to identify GQC-05 as a compound with sufficient selectivity for the c-MYC G-quadruplex to be used in subsequent studies. Three compounds (quindoline-i, NSC176327, and NSC86374)



ΔT_m (°C) as determined by FRET melt, of ten putative G-quadruplex-interactive compounds

NSC number	Compound class[1]	Compound	10 μM	1 μM	0.1 μM
311150	E	GQC-01	19	1	0
651556	E	GQC-02	0	1	1
649091	E	GQC-03	7	0	1
79486	E	GQC-04	0	1	0
338258	E	GQC-05	21	5	1
100594	E	GQC-06	2	1	1
24037	NE	GQC-07	1	1	1
65238	NE	GQC-08	0	0	0
118176	NE	GQC-09	17	1	0
86372	NE	GQC-10	12	3	1

[1]E, ellipticine derivative; NE, non-ellipticine derivative.

Fig. 7 Flow chart showing the procedure utilized to identify GQC-05 as a c-MYC G-quadruplex-interactive compound with enough selectivity to use in subsequent experiments. Initially quindoline-i, NSC176327, and NSC86374 were used in an energy-minimized overlay to give rise to the pharmacophore query. This was used as described in the text to identify NSC338258 (GQC-05) as the compound to use in the subsequent exon-specific CA46 assay and ChIP analysis

were chosen for in silico superimposition based on their reported ability to lower c-MYC transcription or stabilize the c-MYC G-quadruplex. From the energy-minimized overlay, a pharmacophore query was generated, which was used in the NCI and ChemBridge databases to select ten additional compounds for testing in a

fluorescence resonance energy transfer (FRET) melt experiment. Two drug-like property filters, MW and polar surface area, as well as synthetic accessibility were used in this selection. The FRET assay identified NSC338258 (GQC-05) as the compound that gave the maximum change in melting temperature (ΔT_m).

3.2 Validation of GQC-05 as a c-MYC G-Quadruplex-Binding Compound

Once GQC-05 had been selected to move forward, circular dichroism (CD), competition dialysis, and surface plasmon resonance were used to determine the stoichiometry, binding parameters, and selectivity for the c-MYC G-quadruplex vs duplex and single stranded DNA as well as between different G-quadruplexes. CD demonstrated a 2:1 drug:G-quadruplex binding ratio (Fig. 8a), while the K_Ds for the two binding sites were 0.1 $\pm$ 0.01 and 1.14 $\pm$ 0.025 μM, which are about ten times stronger than quindoline-i. Overall, GQC-05 demonstrated more rapid binding and slower dissociation than quindoline-i (Fig. 8b). Finally, competition dialysis showed that GQC-05 had a 45-fold higher preference for the c-MYC G-quadruplex over its duplex sequence and bound with greater preference to the c-MYC G-quadruplex than to other G-quadruplexes or single-stranded and duplex DNA (Fig. 8c).

3.3 Demonstration in a Cellular System That the Downstream Effects of the G-Quadruplex-Interactive Compounds Are Mediated Directly Through the c-MYC G-Quadruplex and Not Indirectly Through Other G-Quadruplexes or Other Cellular Targets

While c-MYC transcription in the translocated allele in the CA46 lymphoma cell line is not under the control of the NHE III_1 and its G-quadruplex, this cell line does have the advantage that its non-translocated allele maintains the integrity of the NHE III_1. Therefore, if ligands that bind to the G-quadruplex do modulate gene expression, their effect in the non-translocated allele relative to the translocated allele can be assessed. The challenge is to find a way to measure independently the c-MYC transcription mediated through the non-translocated (NT) vs translocated (T) allele. Fortunately, this can be done because exon 1 is lost from the T allele. Figure 9a shows the principle of this assay. For a ligand to be considered a specific modulator of c-MYC gene expression through the G-quadruplex, the effect should be seen exclusively through exon 1 (NT), which retains the G-quadruplex, but has little or no effect through exon 2 (T). This is exactly the result we observe for GQC-05 (Fig. 9b), where only in the NT allele do we see an effect on c-MYC transcription.

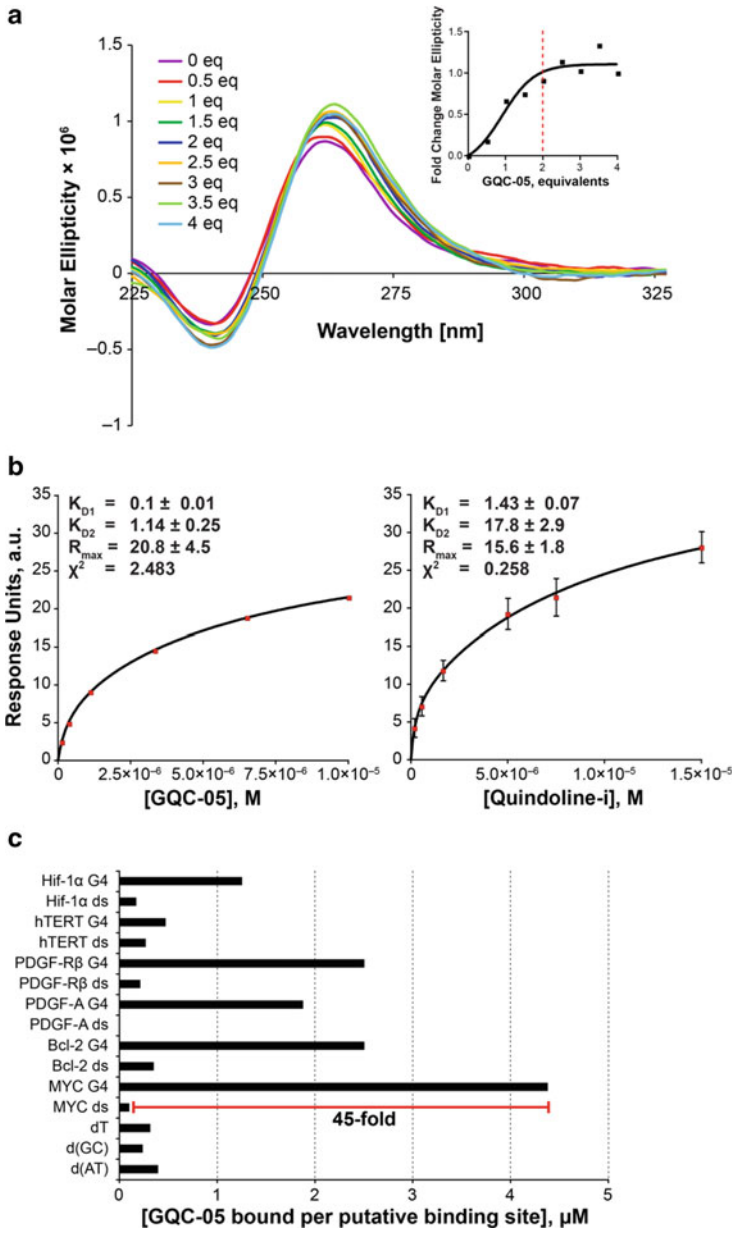


Fig. 8 Interaction and selectivity of GQC-05. (**a**) CD spectra with increasing equivalents of compound demonstrate an increase in the parallel G-quadruplex structure with a peak at 262 nm. *Inset*: quantitation of fraction change in molecular ellipticity at 262 nm as a function of equivalents highlights a plateau at 2 eq. (**b**) GQC-05 demonstrated both more rapid binding and slower dissociation, with associated lower K_D values for both the first and second binding sites, as compared to quindoline-i. Data represent a minimum of triplicate experiments.

Interestingly, quindoline-i did not show an exon-specific effect. These results obtained from GQC-05 are consistent with an effect mediated through the NHE III_1. To further pinpoint the target to the element that contains the G-quadruplex, we performed chromatin immunoprecipitation (ChIP) assays.

3.4 Demonstration That the Activating Transcription Factors Were Displaced from the NHE III_1 as a Consequence of Drug Stabilization of the c-MYC G-Quadruplex

To evaluate further the specificity of the effect of GQC-05 on c-MYC transcription through the G-quadruplex in the NHE III_1, we questioned whether the transcriptional factors associated with the transcriptionally active forms and the dynamic interconversion of the G-quadruplex/i-motif to duplex DNA are displaced in the NT allele in the CA46 cell line. The dynamics of the NHE III_1 region within the c-MYC promoter are shown in Fig. 10a. Transcriptional factors involved in this dynamic equilibrium have been identified through a series of extensive studies [18, 40–42]. ChIP analysis with antibodies specific to the transcriptional factors hnRNP K, CNBP, and Sp1 showed that they are all significantly displaced by 12 h after incubation with GQC-05, and even at 6 h, hnRNP K is significantly reduced (Fig. 10b). In an important control experiment, no significant effects were seen on the expression of these transcriptional factors at 12 h [17]. Thus, the combined results from the CA46 exon-specific assay and the ChIP analysis provide convincing evidence that GQC-05 mediates its effects through the G-quadruplex in the NHE III_1.

4 Molecular Modeling of GQC-05 and Quarfloxin Based on the Quindoline-i:c-MYC G-Quadruplex Complex

While quindoline-i did not show an exon-specific effect, the NMR solution structure of the quindoline-i:c-MYC G-quadruplex complex potentially provides a platform for understanding the binding interactions of other small molecules with the c-MYC G-quadruplex, such as GQC-05 and Quarfloxin [43]. Quarfloxin is a first-in-class G-quadruplex-interactive drug that was advanced to phase 2 clinical trials, and it has been proposed that its antitumor effect is mediated via displacement of nucleolin and subsequent binding of this protein to the c-MYC

Fig. 8 (continued) (**c**) Competition dialysis highlights that GQC-05 binds to the MYC G4 (G-quadruplex) with the highest affinity per binding site, at 45-fold selectivity over the MYC dsDNA (double-stranded DNA) sequence. Figure is modified, with permission, from [17], © (2011) American Society for Biochemistry and Molecular Biology

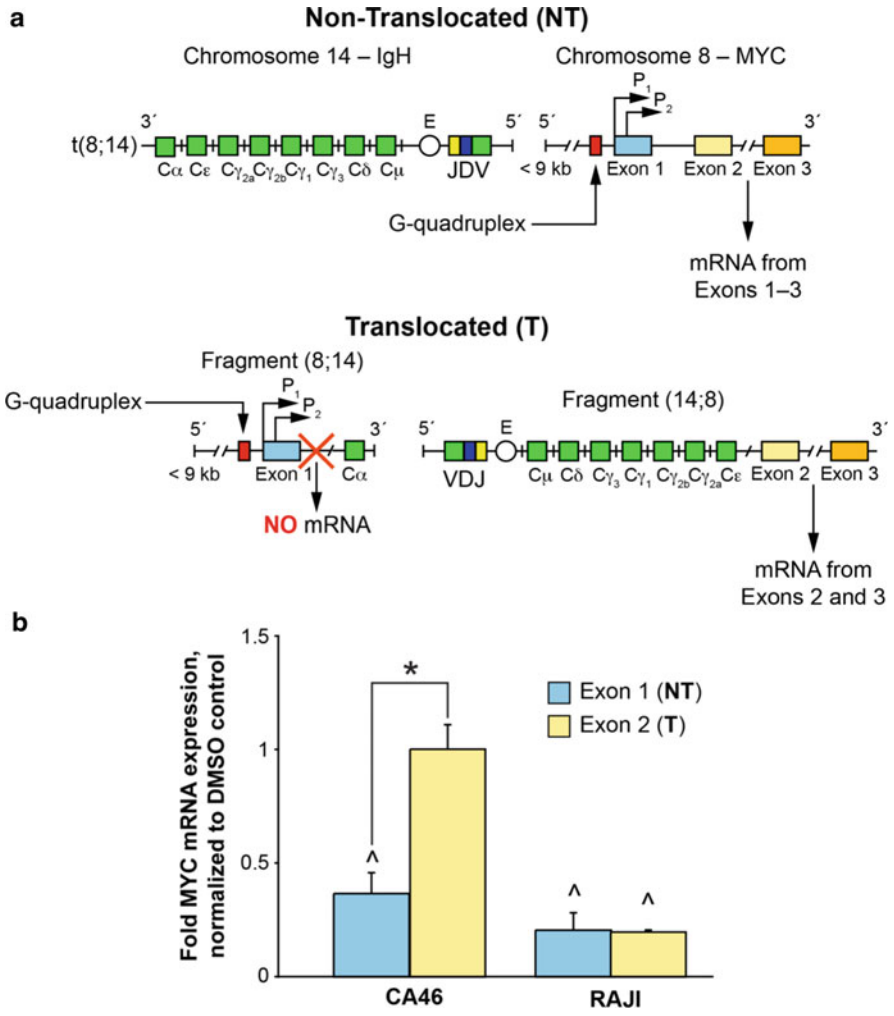


Fig. 9 The exon-specific effect of GQC-05. (**a**) Due to the reciprocal translocation between chromosomes 8 and 14, there are varying resultant MYC mRNAs produced. The NT products are normal, with a functional MYC under the control of a G-quadruplex, whereas the functional MYC produced from the fragment (14;8) on the T allele lacks G-quadruplex-mediated control. The G-quadruplex was removed, along with exon 1, and produces no known product from the fragment (8;14). Measurements of mRNAs containing exon 1 will mirror the NT allele; mRNAs containing exon 2 will show both the T and the NT products. (**b**) GQC-05 decreases MYC mRNA only from the NT allele in the CA46 cells where the G-quadruplex is maintained and has no effect from the T allele. In comparison, there is no exon-specific effect in the RAJI cells where both exons are produced under the mediated control of a G-quadruplex. All mRNA products were normalized to DMSO vehicle control; experiments at each time point are in triplicate; * $p < 0.05$ between exons; ^ $p < 0.05$ as compared to DMSO vehicle controls. Figure is modified, with permission, from [17], © (2011) American Society for Biochemistry and Molecular Biology

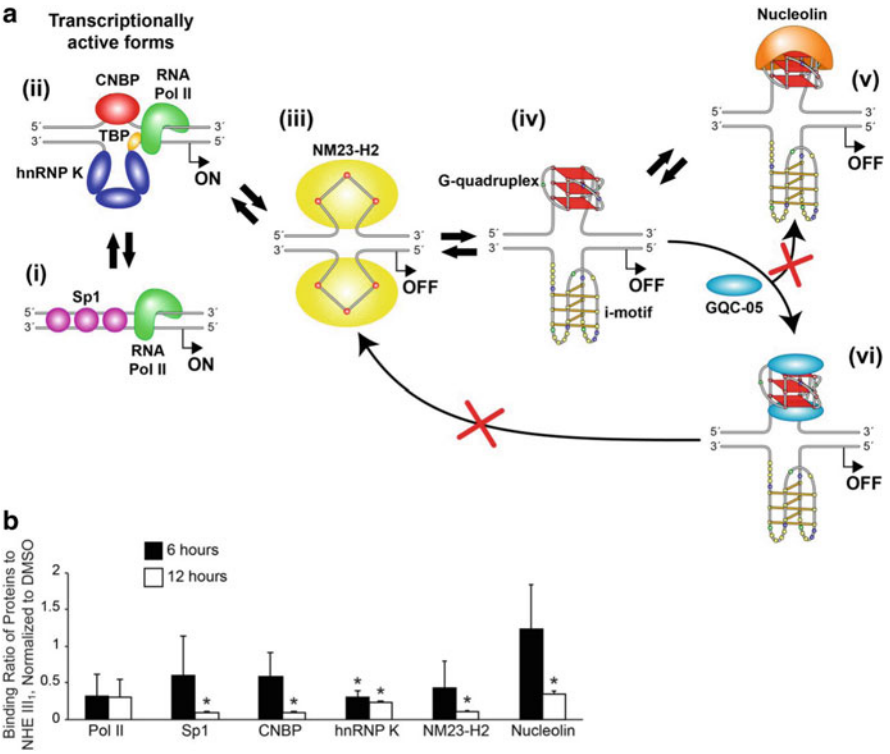


Fig. 10 (**a**) Dynamics of the NHE III_1 region within the MYC promoter. (*i*) Double-stranded DNA is bound by Sp1 in the transcriptionally active form, (*ii*) negative supercoiling opens the region to form single-stranded DNA to which CNBP and hnRNP K can bind to activate MYC. Alternatively (*iv*), the G-quadruplex and i-motif can form within this region, which can be unfolded by NM23-H2 (*iii*). In the absence of GQC-05, (*v*) nucleolin will cap and stabilize the G-quadruplex; however, the binding of two molecules of GQC-05 (*vi*) prevents this capping and stabilizes the MYC G-quadruplex, leading to transcriptional down-regulation. (**b**) ChIP analysis revealed a dynamic change in protein binding to the NHE III_1 region within 6–12 h post-GQC-05 treatment. The binding of hnRNP K was significantly decreased within 6 h, and with the exception of RNA Pol II, which does not change, the binding of all other proteins is also decreased by 12 h. Data represent duplicate ChIP experiments; * $p < 0.05$ as compared to DMSO vehicle controls. Figure is modified, with permission, from [17], © (2011) American Society for Biochemistry and Molecular Biology

G-quadruplex [19]. The in silico docking of these molecules with the c-MYC G-quadruplex may provide insight into the factors governing intermolecular binding interactions.

4.1 Modeling Methods

The solution structure of quindoline-i with the c-MYC G-quadruplex (PDB code: 2L7V) was used as a starting model for docking studies [24]. The model for the wild-type c-MYC sequence structure was generated from this NMR structure using

Table 1 Interaction energy values for small molecules with the c-MYC G-quadruplex wild-type structure

Compound	Interaction energy (kcal/mol)
Quindoline-i	−58.8
GQC-05	−89.7
Quarfloxin	−112.5

Insight II modeling software (Accelrys Inc., San Diego). The wild-type sequence was used throughout the modeling studies, and all charges were assigned using the consistent valence force field. Initial docking orientations were generated by a three-dimensional overlay of quindoline-i with GQC-05 and Quarfloxin. The 5′-end of the c-MYC G-quadruplex was used as the active site for small molecules. The small-molecule:c-MYC G-quadruplex complex was soaked in a 10-Å layer of TIP3P water. The entire assembly was minimized using 100,000 steps of Discover 3.0. This minimization was followed by molecular dynamics involving equilibration of 50 ps and simulations of 450 ps. Frames were collected every picosecond during the simulation phase. All the trajectories were analyzed using potential energy, and the 20 lowest potential energy frames were used to create an average structure. This average structure was then refined using 100,000 steps of minimization. This refined structure was used for calculation of interaction energy values (Table 1).

4.2 Comparison of Binding Interactions

Docking results clearly show differences in the driving forces that determine the binding of these molecules with the c-MYC G-quadruplex. In the case of quindoline-i, the protonated nitrogen atom aligns itself with the central K^+ channel while the quindoline ring interacts with the guanines of the G-tetrad via stacking. Thus, the diethylaminoethyl side chain of quindoline-i is placed in the two-base (5′-GA) double-chain-reversal loop and interacts with the phosphate backbone. As evident from the energy values, Quarfloxin exhibits a better interaction with the c-MYC G-quadruplex than quindoline-i. Quarfloxin, a fluoroquinolone analog, exhibits stacking interactions with an extended aromatic ring system intercalating between the top guanine tetrad and the 5′-end guanine, a site that is created by the movement of the 5′-end adenine. Additionally, the two side chains extend into the two single-base double-chain-reversal loops of the G-quadruplex structure (Fig. 11). Ellipticines are known to intercalate preferentially with GC-rich regions of DNA; the compound GQC-05 is a C9-substituted ellipticine analog. The pyridocarbazole ring of GQC-05 docks in an orientation similar to quindoline-i. However, the interaction of the pyridocarbazole ring with the G-tetrad is governed by stacking interactions, whereas the dimethylaminoethyloxy group in the side chain interacts with the negatively charged phosphate backbone in the thymine 19 single-base double-chain-reversal loop (Fig. 12).

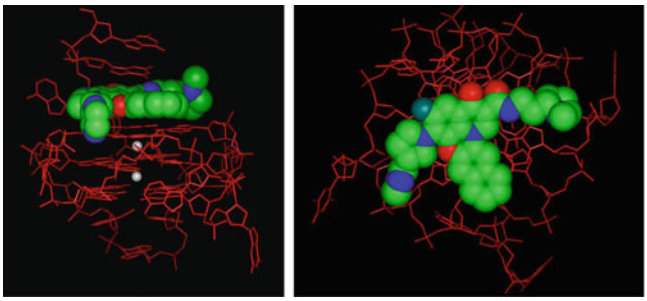


Fig. 11 Molecular models of Quarfloxin with the c-MYC G-quadruplex structure, showing the intercalation between the top tetrad and the 5′-end flanking bases (*left*) and a view looking down on the top tetrad (*right*). Quarfloxin is shown as a space-filled model colored by atom type

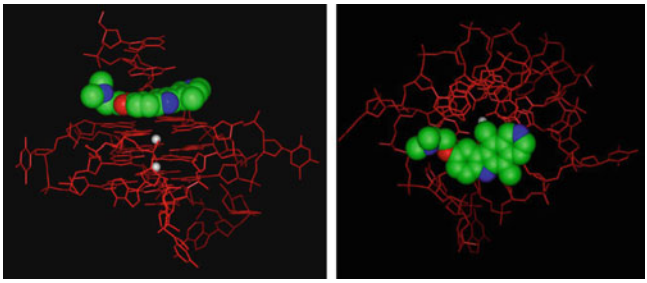


Fig. 12 Molecular model of GQC-05 with the c-MYC G-quadruplex structure, showing the intercalation between the top tetrad and the 5′-end flanking bases (*left*) and a view looking down on the top tetrad (*right*). GQC-05 is shown as a space-filled model colored by atom type

5 Conclusions

We used NMR modeling and biological characterization to target the G-quadruplex in the c-MYC promoter in an attempt to find compounds that could down-regulate the expression of c-MYC at the transcriptional level. Our NMR work indicated a large conformational change in the flanking region of the c-MYC G-quadruplex, resulting in the formation of an induced intercalated triad pocket from the interaction between the G-quadruplex in the c-MYC promoter and the asymmetric small molecule quindoline-i. In our biological investigations, we identified GQC-05 as a potent inhibitor of c-MYC transcription. Subsequent studies using an exon-specific assay demonstrated that the effect on c-MYC transcription is mediated directly through the G-quadruplex in the NHE III_1 of the c-MYC promoter. Finally, we used the NMR structure generated for the 2:1 quindoline-i:c-MYC G-quadruplex complex to glean insights about potential modes of binding for GQC-05 and Quarfloxin to the c-MYC G-quadruplex.

Acknowledgements This research has been supported by grants from the National Institutes of Health (CA153821 [LHH] and 1S10 RR16659 and CA122952 [DY]) and the National Foundation for Cancer Research (VONHOFF0601 [LHH]). We are grateful to Dr. David Bishop for preparation, proofreading, and editing of the final version of the manuscript.

References

1. Kohn KW (1996) Beyond DNA cross-linking: history and prospects of DNA-targeted cancer treatment—fifteenth Bruce F. Cain Memorial Award Lecture. Cancer Res 56:5533–5546
2. Balasubramanian S, Hurley LH, Neidle S (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? Nat Rev Drug Discov 10:261–275
3. Huppert JL, Balasubramanian S (2007) G-quadruplexes in promoters throughout the human genome. Nucleic Acids Res 35:406–413
4. Sen D, Gilbert W (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. Nature 334:364–366
5. Sundquist WI, Klug A (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. Nature 342:825–829
6. Zahler AM, Williamson JR, Cech TR, Prescott DM (1991) Inhibition of telomerase by G-quartet DNA structures. Nature 350:718–720
7. Sun D, Thompson B, Cathers BE, Salazar M, Kerwin SM, Trent JO, Jenkins TC, Neidle S, Hurley LH (1997) Inhibition of human telomerase by a G-quadruplex-interactive compound. J Med Chem 40:2113–2116
8. Huppert JL, Balasubramanian S (2005) Prevalence of quadruplexes in the human genome. Nucleic Acids Res 33:2908–2916
9. Kikin O, D'Antonio L, Bagga PS (2006) QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. Nucleic Acids Res 34:W676–W682
10. Todd AK, Johnston M, Neidle S (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. Nucleic Acids Res 33:2901–2907
11. Huppert JL, Bugaut A, Kumari S, Balasubramanian S (2008) G-quadruplexes: the beginning and end of UTRs. Nucleic Acids Res 36:6260–6268
12. Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, Lander ES, Kellis M (2005) Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. Nature 434:338–345
13. Kouzine F, Sanford S, Elisha-Feil Z, Levens D (2008) The functional response of upstream DNA to dynamic supercoiling *in vivo*. Nat Struct Mol Biol 15:146–154
14. Sun D, Hurley LH (2009) The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-Myc promoter: implications for drug targeting and control of gene expression. J Med Chem 52:2863–2874
15. Brooks TA, Kendrick S, Hurley L (2010) Making sense of G-quadruplex and i-motif functions in oncogene promoters. FEBS J 277:3459–3469
16. Yu Z, Gaerig V, Cui Y, Kang H, Gokhale V, Zhao Y, Hurley LH, Mao H (2012) Tertiary DNA structure in the single-stranded hTERT promoter fragment unfolds and refolds by parallel pathways via cooperative or sequential events. J Am Chem Soc 134:5157–5164
17. Brown RV, Danford FL, Gokhale V, Hurley LH, Brooks TA (2011) Demonstration that drug-targeted down-regulation of MYC in non-Hodgkins lymphoma is directly mediated through the promoter G-quadruplex. J Biol Chem 286:41018–41027
18. Brooks TA, Hurley LH (2009) The role of supercoiling in transcriptional control of MYC and its importance in molecular therapeutics. Nat Rev Cancer 9:849–861
19. Brooks TA, Hurley LH (2010) Targeting MYC expression through G-quadruplexes. Genes Cancer 1:641–649

20. Michelotti GA, Michelotti EF, Pullner A, Duncan RC, Eick D, Levens D (1996) Multiple single-stranded *cis* elements are associated with activated chromatin of the human c-*myc* gene in vivo. Mol Cell Biol 16:2656–2669
21. Mathur V, Verma A, Maiti S, Chowdhury S (2004) Thermodynamics of *i*-tetraplex formation in the nuclease hypersensitive element of human *c-myc* promoter. Biochem Biophys Res Commun 320:1220–1227
22. Simonsson T, Pecinka P, Kubista M (1998) DNA tetraplex formation in the control region of c-*myc*. Nucleic Acids Res 26:1167–1172
23. Soucek L, Whitfield J, Martins CP, Finch AJ, Murphy DJ, Sodir NM, Karnezis AN, Swigart LB, Nasi S, Evan GI (2008) Modelling myc inhibition as a cancer therapy. Nature 455:679–683
24. Dai J, Carver M, Hurley LH, Yang D (2011) Solution structure of a 2:1 quindoline–c-MYC G-quadruplex: insights into G-quadruplex-interactive small molecule drug design. J Am Chem Soc 133:17673–17680
25. Ambrus A, Chen D, Dai J, Jones RA, Yang D (2005) Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. Biochemistry 44:2048–2058
26. Yang D, Okamoto K (2010) Structural insights into G-quadruplexes: towards new anticancer drugs. Future Med Chem 2:619–646
27. Guyen B, Schultes CM, Hazel P, Mann J, Neidle S (2004) Synthesis and evaluation of analogues of 10*H*-indolo[3,2-*b*]quinoline as G-quadruplex stabilising ligands and potential inhibitors of the enzyme telomerase. Org Biomol Chem 2:981–988
28. Zhou J-L, Lu Y-J, Ou T-M, Zhou J-M, Huang Z-S, Zhu X-F, Du C-J, Bu X-Z, Ma L, Gu L-Q, Li Y-M, Chan AS-C (2005) Synthesis and evaluation of quindoline derivatives as G-quadruplex inducing and stabilizing ligands and potential inhibitors of telomerase. J Med Chem 48:7315–7321
29. Ou TM, Lu YJ, Zhang C, Huang ZS, Wang XD, Tan JH, Chen Y, Ma DL, Wong KY, Tang JC, Chan AS, Gu LQ (2007) Stabilization of G-quadruplex DNA and down-regulation of oncogene c-myc by quindoline derivatives. J Med Chem 50:1465–1474
30. Mandal M, Breaker RR (2004) Gene regulation by riboswitches. Nat Rev Mol Cell Biol 5:451–463
31. Shin-ya K, Wierzba K, Matsuo K, Ohtani T, Yamada Y, Furihata K, Hayakawa Y, Seto H (2001) Telomestatin, a novel telomerase inhibitor from *Streptomyces anulatus*. J Am Chem Soc 123:1262–1263
32. Han FX, Wheelhouse RT, Hurley LH (1999) Interactions of TMPyP4 and TMPyP2 with quadruplex DNA. Structural basis for the differential effects on telomerase inhibition. J Am Chem Soc 121:3561–3570
33. Neidle S (2009) The structures of quadruplex nucleic acids and their drug complexes. Curr Opin Struct Biol 19:239–250
34. Phan AT, Kuryavyi V, Gaw HY, Patel DJ (2005) Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. Nat Chem Biol 1:167–173
35. Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-*MYC* transcription. Proc Natl Acad Sci USA 99:11593–11598
36. Campbell NH, Parkinson GN, Reszka AP, Neidle S (2008) Structural basis of DNA quadruplex recognition by an acridine drug. J Am Chem Soc 130:6722–6724
37. Haider SM, Parkinson GN, Neidle S (2003) Structure of a G-quadruplex–ligand complex. J Mol Biol 326:117–125
38. Qin Y, Hurley LH (2008) Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. Biochimie 90:1149–1171
39. Kang HJ, Park HJ (2009) Novel molecular mechanism for actinomycin D activity as an oncogenic promoter G-quadruplex binder. Biochemistry 48:7392–7398

40. Dexheimer TS, Carey SS, Zuohe S, Gokhale VM, Hu X, Murata LB, Maes EM, Weichsel A, Sun D, Meuillet EJ, Montfort WR, Hurley LH (2009) NM23-H2 may play an indirect role in transcriptional activation of c-myc gene expression but does not cleave the nuclease hypersensitive element III_1. Mol Cancer Ther 8:1363–1377
41. González V, Hurley LH (2010) The c-*MYC* NHE III_1: function and regulation. Annu Rev Pharmacol Toxicol 50:111–129
42. González V, Hurley LH (2010) The C-terminus of nucleolin promotes the formation of the c-*MYC* G-quadruplex and inhibits c-*MYC* promoter activity. Biochemistry 49:9706–9714
43. Drygin D, Siddiqui-Jain A, O'Brien S, Schwaebe M, Lin A, Bliesath J, Ho CB, Proffitt C, Trent K, Whitten JP, Lim JK, Von Hoff D, Anderes K, Rice WG (2009) Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis. Cancer Res 69:7653–7661

Top Curr Chem (2013) 330: 23–46
DOI: 10.1007/128_2012_350

Published online: 12 July 2012

Higher-Order Quadruplex Structures

Luigi Petraccone

Abstract Structural studies have shown that four G-tracts along a DNA strand are the minimal requirement for intramolecular G-quadruplex formation. Longer DNA sequences containing multiples of four G-tracts could, in principle, form higher-order structures based on multiple G-quadruplex blocks. This latter condition is abundantly verified for the telomeric single-stranded overhang (~200 nt) consisting of tens of TTAGGG repeats, thus opening new interesting questions about the structure of the "real" telomeric DNA. How many quadruplex units form in the human telomeric overhang? Which type of quadruplex topologies? Do they interact or not? What about their binding properties? Although many of these questions are still unanswered, recent experimental and computational studies have begun to address them. The existence and relevance of these higher-order quadruplex structures in the human genome is now an interesting and stimulating research topic in the quadruplex field. The recent results, the unsolved problems, and the future prospects for understanding higher-order telomeric quadruplex structures are the main topics of this review. Other studies on long telomeric RNA sequences and on other intramolecular (non telomeric) DNA higher order quadruplex structures are also presented.

Keywords G-quadruplex · Higher order structure · Molecular modeling · Telomeric DNA · Telomeric RNA · Thermodynamic stability

Contents

1 Introduction 24
2 Experimental Evidence of Multi-quadruplex Structures in Human Telomeric DNA 26
3 Toward a Molecular Model of Long Human Telomeric Sequences 28

L. Petraccone (✉)
Dipartimento di Scienze Chimiche, Università degli Studi di Napoli "Federico II", Via Cintia, 80126 Napoli, Italy
e-mail: luigi.petraccone@unina.it

3.1 Parallel-Based Higher Order Quadruplex Structures 29
3.2 Hybrid-Based Higher Order Quadruplex Structures 30
4 How to Discriminate Between Alternative "Plausible" Higher Order Quadruplex Models? 32
5 Energetics of Higher-Order Human Telomeric Quadruplexes Formation 35
6 Binding Properties of Higher Order Quadruplex Structures: New Prospects for Drug Design? 38
7 What About Long RNA Telomeric Sequences? 39
8 Non-telomeric Higher-Order Quadruplex Structures 40
9 Concluding Remarks 42
References 43

1 Introduction

Single quadruplex units are building blocks for higher-order quadruplex structures both intramolecular and intermolecular, depending on whether or not the quadruplex blocks belong to the same DNA (or RNA) strand. In the last 5 years, intramolecular higher-order quadruplex structures have been the subject of increasing interest due to their relevance as a model for the full length telomeric DNA structure [1–9]. Indeed, structural studies have shown that four G-tracts along a telomeric DNA strand are the minimal requirement for intramolecular G-quadruplex formation but that the telomeric single-stranded overhang consists of tens of TTAGGG repeats (~200 nt) and may form higher order structures involving multiple quadruplex units. As telomeric DNA is a promising anti-cancer drugs target, detailed knowledge of its physiological relevant structure is of paramount importance for structural-based drug design.

Despite this, so far structural studies on human telomeric DNA have been limited to short DNA sequences (no more than four G-tracts) able to form individual quadruplexes (Fig. 1a). The first topology for a human telomeric quadruplex sequence to be reported in a K^+ environment was determined by crystallographic analysis [10]. This structure – called "parallel" or "propeller" – has all four G-tracts parallel and connected by three propeller loops. Later, two distinct (3+1) topologies, termed hybrid-1 and hybrid-2, have been reported using NMR methods in solution [11–13]. Both structures have three G-tracts oriented in one direction and the fourth in the opposite direction, but differ in the order of the loop arrangements. These two forms appear to coexist in K^+ solution and the equilibrium between them is governed by the presence of 3′-flanking bases [6]. Recently, an antiparallel basket type form in K^+ solution – called "Form-3" – have been reported [14]. This is distinct from the known antiparallel topology in Na^+ solution [15] and contains just two G-quartets. This structural polymorphism has stimulated wide debate: which of these topologies is the most biologically relevant? Which is the prevalent quadruplex structure in physiological conditions?

Although there is not yet a final answer, it is clear from these structural studies that the particular truncation of the full length telomeric DNA employed dictates the structure that is observed. Several four-repeat human telomeric sequences which start with a G (i.e., there are no flaking bases at the 5′-extremity) adopt

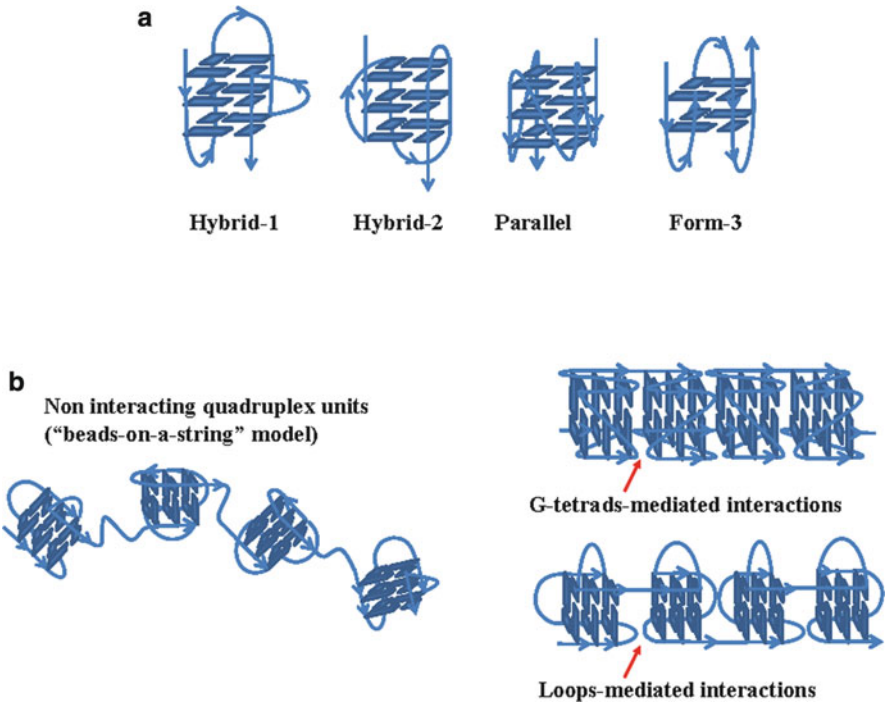


Fig. 1 (**a**) Schematic structures of the single quadruplex topologies (hybrid-1, hybrid-2, Form-3 and Parallel) determined for the four-repeat human telomeric sequences in K^+ solution (the *arrows* are in the 3′–5′ direction). (**b**) Possible structures for long telomeric sequences based on the single quadruplexes as building blocks. The quadruplex subunits could be not interacting (beads-on-a-string model) or could interact through staking interactions involving directly the G-tetrads cores or the loop bases

form-3 in K^+ solution [14]. Sequences with both 5′-flaking and 3′-flaking sequences adopt the hybrid conformation; hybrid-2 is favored by the presence of 3′-flaking bases [6]. What about the "real" telomeric overhang whereby each four-repeat unit has tens of TTAGGG repeats at both its extremities? When considering a quadruplex formed inside the full length telomeric DNA, the presence of flanking bases seems to be more realistic in mimicking the physiological condition. However, the possibility of forming consecutive quadruplex blocks adds another dimension to this already complex problem, while opening new interesting questions about the structure of the "real" telomeric DNA (Fig. 1b). How many quadruplex units form in the human telomeric overhang? What types of quadruplex topologies exist? Do the contiguous quadruplex units interact or not? What about their binding properties? Although many of these questions are still unanswered, recent experimental and computational studies have begun to address them. The question about the existence and relevance of these higher-order quadruplex structures in the human telomere is now an interesting and stimulating research topic in the quadruplex field. The recent results, the unsolved problems, and the future

prospects for understanding higher-order telomeric quadruplex structures are the main topics of this review. Other studies on long telomeric RNA sequences and on other intramolecular (non telomeric) DNA higher order quadruplex structures are also presented.

2 Experimental Evidence of Multi-quadruplex Structures in Human Telomeric DNA

In the last 5 years the properties of long telomeric sequences have been the subject of several experimental studies. Telomeric sequences with 8, 12, and 16 repeats, that potentially can form 2, 3, and 4 consecutive quadruplexes, have been employed (Table 1). What experimental evidence exists to show that these sequences really form intramolecular structures based on the single quadruplexes as building blocks? Data collected with different methodologies strongly point to this direction. Gel electrophoresis and analytical ultracentrifugation (AUC) data demonstrated that longer telomeric sequences form mainly compact intramolecular structures consistent with the formation of contiguous quadruplex units [1, 5, 8, 20, 21]. A further indication of the multi-quadruplex nature of these structures has been inferred from the comparison of their circular dichroism (CD) spectra with the CD spectrum of the standard single quadruplex forming sequences [such as $(TTAGGG)_4$ or slight variants thereof]. It has been observed that the CD spectrum of sequences formed by integral multiples of the four G-tracts [such as $(TTAGGG)_n$ with $n = 8, 12, 16$] has a similar shape to the CD spectrum of the single quadruplex but with an intensity that approximately corresponds to an integral multiple of the monomer intensity [1, 2, 5]. The simplest interpretation of this observation leads to the multi-quadruplex structure of the longer sequences: another quadruplex unit is added to the previous one when the DNA sequence is elongated.

Other indirect evidence of the existence of these multi-quadruplex structures comes from the analysis of their melting behavior. Van't Hoff analysis of the UV melting process for the $(TTAGGG)_n$ with $n = 8$ and 12 revealed a ΔH° values close to the value expected for a quadruplex melting (~200 kJ/mol). This observation suggests that the $(TTAGGG)_8$ and $(TTAGGG)_{12}$ structures are divided into more cooperative units which melt independently, each unit having (approximately) the "size" of a quadruplex [1, 22]. A similar result has been obtained for the eight-repeat sequence in the presence of 40% (w/v) PEG200 [23]. On the other hand, the total enthalpy values for the melting of 8-repeat and 12-repeat sequences has been measured directly by means of differential scanning calorimetry (DSC) yielding values more consistent with the presence of 2 and 3 quadruplex units, respectively [5]. The thermodynamics of these higher-order structures' melting will be discussed in more detail later in the chapter (see Sect. 5).

Several experimental studies were done on the eight-repeat sequences [such as $(TTAGGG)_8$ or slight variants thereof] [2, 9, 16]. In one of the early studies, the ability

Table 1 Studied DNA telomeric sequences with an integral multiple of the four G_3 blocks

		Solution conditions			
Telomeric sequences	N°[a]	Conc. (mM)	Salt	Subunit structure[b]	Reference
$G_3(T_2AG_3)_7$	n.d.[e]	100	KCl	Various[c]	[8]
$G_3(T_2AG_3)_7$	1–2	2.5	KCl	Hybrid	[9]
$G_3(T_2AG_3)_7$	2	100.0	NH_4OAc	Hybrid	[16]
$AG_3(T_2AG_3)_7$	2	100.0	KCl	Hybrid	[4]
$AG_3(T_2AG_3)_7$	2	10	KCl	Hybrid	[17]
		90.0	LiCl		
$(T_2AG_3)_8$	2	100.0	KCl	Hybrid[d]	[1]
$(T_2AG_3)_8$	2	100.0	KCl	Hybrid	[5]
$(T_2AG_3)_8$	1–2	50	KCl	n.d.	[18]
		10	$MgCl_2$		
$(T_2AG_3)_8\ T_2$	2	100	KCl	Hybrid	[2, 5, 19]
$G_3(T_2AG_3)_{11}$	1–3	2.5	KCl	Hybrid	[9]
$G_3(T_2AG_3)_{11}$	n.d.	100	KCl	Various[c]	[8]
$AG_3(T_2AG_3)_7$	2–3	10	KCl	Hybrid	[17]
		90	LiCl		
$(T_2AG_3)_{12}$	3	100	KCl	Hybrid[d]	[1]
$(T_2AG_3)_{12}$	3	100	KCl	Hybrid	[5]
$(T_2AG_3)_8\ T_2$	3	100	KCl	Hybrid	[5]
$G_3(T_2AG_3)_{15}$	n.d.	100	KCl	Various[c]	[8]
$(T_2AG_3)_{16}$	4	100	KCl	n.d.	[4]
		10	$MgCl_2$		
$(T_2AG_3)_{16}$	2	500	KCl	n.d.	[18]
		10	$MgCl_2$		

[a]Number of quadruplex subunits; more than one integral number indicates that a distribution has been observed (i.e., 2–3 means that DNA strands containing 2 or 3 quadruplex subunits were observed)
[b]Suggested quadruplex structure as inferred from the CD spectrum
[c]The quadruplex conformation was found to be dependent on DNA concentration (going in the direction antiparallel – hybrid – parallel on increasing DNA concentration)
[d]In the original paper, the authors suggested a mix of parallel and antiparallel as the hybrid structure and its CD spectrum was unknown
[e]n.d. stands for "not determined"

of the eight-repeat telomeric sequence to form a two-quadruplex structure has been demonstrated by means of CD, AUC, steady-state fluorescence measurements and molecular modeling [2]. Later, the formation of two contiguous quadruplexes in the $GGG(TTAGGG)_7$ DNA telomeric sequence has been confirmed on the basis of ESI–MS experiments by counting the number of ammonium ions selectively trapped in the structure [16]. Other authors demonstrated by using FRET that two contiguous quadruplexes are formed in the $AGGG(TTAGGG)_7$ sequence [4].

Another remarkable demonstration of a multi-quadruplex structure was obtained by means of atomic force microscopy (AFM) experiments on the $(TTAGGG)_{16}$ sequence [4]. AFM images showed that four contiguous quadruplexes, separated by only a short single-stranded TTA linker, readily formed in the 96 nt telomeric

repeat sequence, although the resolution of the method could not definitively assign the conformations of the individual quadruplex units.

In contrast to these studies, there are scattered reports on sequences with 8, 12, and 16 repeats that suggest that the number of quadruplex subunits formed in these sequences is lower than the value expected on the basis of their length [9, 17, 18]. For example, it has been reported that significant fractions of a 16-repeat and 12-repeat form only 2 quadruplex subunits (instead of 4 and 3, respectively) [17, 18]. However, the source of this discrepancy is not clear and should be further explored. The lower KCl concentrations employed in these studies could be a key element in explaining the observed results.

3 Toward a Molecular Model of Long Human Telomeric Sequences

Experimental evidence suggests that long human telomeric sequences form intramolecular structures formed by consecutive blocks of single quadruplex-like structures. However, the way from this "qualitative" information to a high-resolution structure of the full human telomeric overhang is long. What is the folding of the individual quadruplex units within the multi-quadruplex structure? How do they interact? Regarding this latter issue, there are two opposite possibilities: (1) the G-quadruplex blocks do not interact and can move relatively independently of each other, constrained only by the connecting TTA linkers (this model is called "beads-on-a-string") [1] or (2) G-quadruplex blocks strongly interact through stacking interactions involving directly the G-tetrads cores or the loop bases (Fig. 1b).

Recently, molecular modeling studies have provided some useful suggestions in addressing this and other questions [3, 5, 19]. The basic idea of these studies is that the most "natural" candidates for building blocks of a multi-quadruplex structures are the known individual quadruplex structures formed by the four-repeat human telomeric sequences. Among these, the hybrid and parallel quadruplexes provide efficient scaffolds for a compact-stacking multi-quadruplex structure of human telomeric DNA. Indeed, the 5′ end and the 3′ end of these quadruplex structures point in opposing directions and additional hybrid-type or propeller quadruplexes can be linked end-to-end to form linear multi-quadruplex structures [6, 24]. Following this point of view, several higher order quadruplex structures have been built by joining hybrid and parallel quadruplex as building blocks (Fig. 2). Molecular modeling techniques have been employed to evaluate which molecular models were stereochemically and energetically "plausible" higher-order quadruplex structures. Although experimental validation of the models is always needed, all-atom molecular dynamics (MD) simulations in explicit solvent have provided interesting information on the dynamics and stability of the quadruplex subunits, highlighting the specific implications of each structural model.

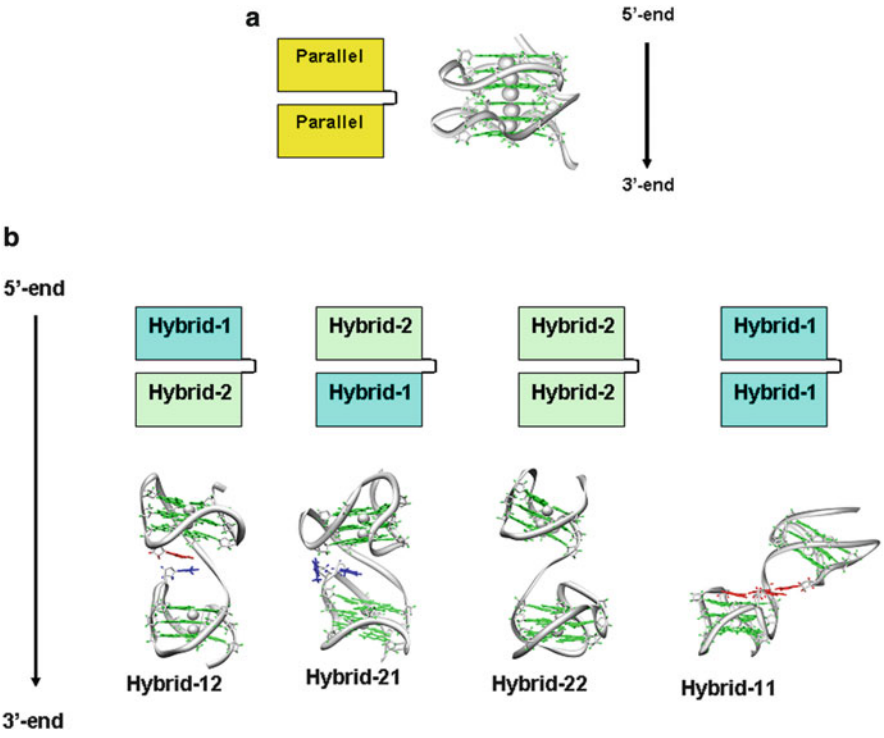


Fig. 2 Schematic representation and computed structures of the parallel (**a**) and hybrid-based (**b**) models of the eight-repeat human telomeric sequence containing two quadruplex subunits. For clarity, in the computed structures loop bases not involved in a quadruplex–quadruplex interaction are not shown. The 5′–3′ orientation is indicated by the *black arrows*. Adapted from [19]

3.1 Parallel-Based Higher Order Quadruplex Structures

A multi-quadruplex structure formed by consecutive parallel quadruplexes was first proposed by Neidle and coworkers [3]. In this model consecutive quadruplexes are connected by a TTA loop in a 5′–3′ orientation and the terminal G-tetrad of each quadruplex subunit stacks over the initial G-tetrad of the subsequent quadruplex subunits, forming extended stacks of such quadruplex structures (Fig. 2a). MD simulations of such molecular structures containing two, three, or four consecutive parallel quadruplexes have been carried out by two groups reaching similar results: the parallel-based higher order quadruplex structures are stable and particularly rigid as expected due to the large stacking interactions existing between each subunits [3, 5, 19]. Flexibility of the models decreases on increasing the number of G-quadruplex subunits (from two to four) due to the greater number of stacking interactions between the G-tetrads. The most flexible part of the model are the loops; however the conformational changes observed in the loops do not have any

impact on the structure of the central G-tetrads core. All the TTA loops in the structural models have similar conformation and are arranged diagonally, external to the guanine tetrads in a propeller-like manner characteristic of the parallel quadruplex topology. The dA nucleotide in each TTA sequence is swung back such that it intercalates between the two T bases and makes stacking interactions with one of the T bases. The contributions of the TTA loops to the total free energy, as estimated by molecular mechanics and Poisson–Boltzmann surface area calculations (MM–PBSA), are very similar [3].

There is no doubt that the parallel quadruplex with its external propeller loops and its exposed terminal G-tetrads seems an ideal topology with which to form extended stacked structures and these MD studies have shown that it can produce stereochemically and energetically plausible higher-order structures.

Nevertheless, experimental evidence of parallel-based higher order quadruplex structures formation is still absent and this model appears inconsistent with the CD spectra in solution reported for longer telomeric sequences (Table 1). Further, experimental AFM images of the $(TTAGGG)_{16}$ sequence clearly show peaks and valleys corresponding to the quadruplex subunits and the TTA linkers connecting them [4], whereas a compact cylindrical structure with contiguous quadruplexes stacked upon one another (without knots) should be expected on the basis of a parallel quadruplex higher-order model [3]. However, recent gel electrophoresis and CD data performed on the eight-repeat sequence in solution with a greatly diminished water activity [in buffers containing 40% (w/v) PEG200] suggest that, in this condition, the parallel-based structure could be slightly favored over other higher order structures (as was previously shown for the four-repeat sequence) [23, 25]. Whether or not, these experimental conditions (of low water activity) are representative of the cellular environment is still matter of debate [26–28].

3.2 Hybrid-Based Higher Order Quadruplex Structures

Various structural models have been built by combining in different ways the hybrid-1 and hybrid-2 conformations to form a two-quadruplex structure [19]. Four possibilities exist in combining the two hybrid-type quadruplex in a 5′–3′ orientation to form a two-quadruplex high-order structure: two consecutive hybrid-1 (Hybrid-11) or hybrid-2 (Hybrid-22) quadruplexes, one hybrid-1 followed by a hybrid-2 (Hybrid-12) or one hybrid-2 followed by a hybrid-1 quadruplex (Hybrid-21). Molecular modeling studies have shown that each of these possibilities results in a quadruplex–quadruplex junction with different structural and dynamic properties (Fig. 2b) [19]. For example, the Hybrid-12 and Hybrid-21 models are not equivalent. In the Hybrid-12 model an interaction between the second loop of the hybrid-1 quadruplex and the second loop of the hybrid-2 quadruplex has been observed, whereas in the Hybrid-21 model there is an interaction between the third loops of both quadruplex units. In the same way, the Hybrid-11 and Hybrid-22 structures are quite different from each other and from the mixed-type structures (Hybrid-12 and

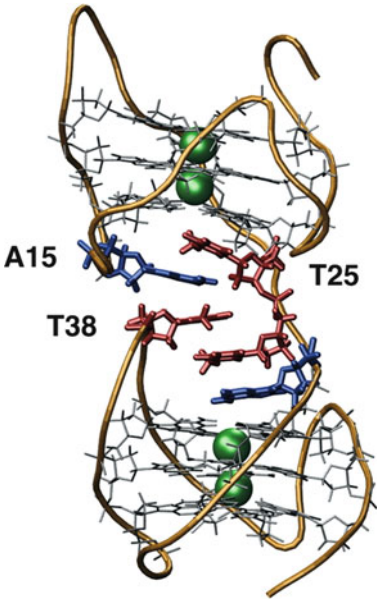


Fig. 3 Average structure from MD simulation of the Hybrid-12 model. The *spheres* are the potassium ions coordinated between the G-tetrads. The loops residues (A15, T25 and T38) involved in the quadruplex–quadruplex interaction are also shown (from [2])

Hybrid-21). Hybrid-11 shows a bending between the quadruplex units stabilized by a specific interaction between dA residues belonging to different subunits. In contrast, the Hybrid-22 model does not show any significant interaction between the quadruplex units. Further, all the hybrid-based structures are clearly different from the parallel model that has extensive quadruplex–quadruplex stacking interactions directly mediated by G-tetrads. The parallel model is the most compact one with an average 5′–3′ distance of 32 Å whereas for all the other models this distance is in the range 42–52 Å. Among all the hybrid-like models, the Hybrid-12 is the one with the most stable linkage between the two quadruplex units within the structure (Fig. 3). In particular, the adenine A15 (belonging to the first quadruplex) forms stable stacking interactions with the thymines T38 and T25 (belonging to the second quadruplex) for 95% of the MD trajectory. Further, it has been shown that, when the MD simulation starts from an initial conformation in which these bases do not interact, the A15-T38 is recovered over the course of the simulation, thus suggesting that this interaction is stable and characteristic of the Hybrid-12 model [2, 19].

As each quadruplex–quadruplex junction is a unique conformation, the molecular surfaces are considerably different, which could have significant implications for structure-based drug design. For instance, Hybrid-11, because of the bending of the two quadruplex units, presents an accessible planar surface corresponding to an A-triad structural motif that is not present in the other models. Despite these structural differences the free energies of these models, as evaluated by MM–PBSA calculations, are very similar, suggesting that the free energy barriers between the different hybrid-based two-quadruplex structures are very small, which is also

consistent with the low free energy barrier reported for the different conformations of the single quadruplex unit folding [29].

What these results bring to light is that, because of the asymmetry of the hybrid-types G-quadruplex structure, four different quadruplex–quadruplex junctions are possible, each characterized by specific shape, electrostatic, and dynamic properties [19]. In this regard it should be noted that a multi-quadruplex structure formed by only one type of hybrid should have only one type of junction (Hybrid-22 or Hybrid-11). The presence of the Hybrid-12 motif in telomeric sequences with more than two quadruplex units implies the presence of either both the Hybrid-12 and Hybrid-21 junctions. Thus, the existence of the mixed-hybrid higher-order quadruplex structure instead of one built from a single type of hybrid structure could provide a better means to modulate protein recognition and regulation of telomere biology. Further, the small free energy barriers between the different conformations (Hybrid-11, Hybrid-22, Hybrid-12, Hybrid-21, and all-parallel) suggests that an equilibrium between different forms could be very sensitive to the local microenvironment as well as to the binding of small molecules to a particular quadruplex–quadruplex junction.

A three-quadruplex structure formed by alternating hybrid-1 and hybrid-2 quadruplex subunits (corresponding to the model 5′-hybrid-1-hybrid-2-hybrid-1-3′, called Hybrid-121) has recently been simulated and proposed as representative of the 12 repeat $(TTAGGG)_{12}$ structure [5]. In this model, there are two different quadruplex junctions (following the DNA strand in the 5′–3′ direction): the hybrid-1-hybrid-2 and the hybrid-2-hybrid-1. The interactions between the two quadruplex units at the hybrid-1-hybrid-2 junction are retained during 10 ns of MD simulation of the Hybrid-121 model as it was observed for the two-quadruplex Hybrid-12 model confirming that this type of quadruplex–quadruplex interaction (QQI) is particularly stable. To date no other models combining hybrid with parallel or antiparallel or other quadruplex folding types have been subject of modeling studies and proposed as "plausible" alternative structures. As will be shown later, the hybrid-based models are, at present, the most consistent with the available data on longer telomeric sequences.

4 How to Discriminate Between Alternative "Plausible" Higher Order Quadruplex Models?

As described above, different higher order quadruplex structures have been proposed for longer DNA telomeric sequences depending on the type of quadruplexes involved and on the kind of interactions between the quadruplex units. All the proposed models are in principle potential candidates to represent the "real" DNA telomeric structures. How is it possible to distinguish these models experimentally? Which one might be most representative of a high-resolution structure of the telomeric DNA? Resolution of this problem is not an easy task for standard

structural techniques like X-ray crystallographic and NMR because of their inherent difficulty of coping with longer DNA sequences.

To address this issue, an innovative approach combining different experimental and computational approaches to discriminate among several plausible structures has been reported [19]. The basic idea is to validate the structural models by means of experimental measurements. To apply the method, one should look at some physico-chemical properties that can be readily computed from the structural models and, at the same time, can be measured experimentally with enough accuracy. The comparison of the predicted and the measured values of these properties can then be employed as test of validity of the structural models. This approach has been pioneered by the groups of Prof. J.B. Chaires and Prof. J.O. Trent (University of Louisville, KY, USA) to show that the parallel conformation was not the dominant structure adopted in solution by the four-repeat human telomeric sequence [30]. They employed the program HYDROPRO [31] to compute from the crystallographic atomic coordinates of the parallel quadruplex the expected value of the sedimentation coefficient and found it significantly different from the value experimentally measured in K^+ solution by means of AUC experiments. The same methodology was later employed to discriminate between two structural models for the three-quadruplex forming sequence $(TTAGGG)_{12}$, one involving all parallel quadruplex and the other containing both hybrid-1 and hybrid-2 quadruplexes (the Hybrid-121 model discussed in Sect. 3.2). The expected sedimentation coefficients distributions were computed from the MD trajectories of these two higher order quadruplex models and were compared with the experimental sedimentation coefficient distribution [5]. The measured sedimentation coefficient has been found to be consistent with the predicted value of the hybrid-based model whereas it has been found to be significantly different from the value predicted on the basis of the parallel-based model (Fig. 4a). On the other hand, the CD spectra of $(TTAGGG)_{12}$ were also characteristic of the hybrid-type structures (Table 1). Although these results do not allow unequivocal proof of the three-quadruplex structure, it clearly suggests that a major fraction of $(TTAGGG)_{12}$ can fold in three contiguous quadruplex units and that these units are most likely a mixture of hybrid conformations. The same results were obtained with the $(TTAGGG)_{12}TT$ sequence, suggesting that the presence of the 3′-flanking bases are less important in determining the quadruplex units folding in comparison with their drastic effect on the folding of the short telomeric sequences [5, 6].

Another useful experimental property able to discriminate between different higher order quadruplex models is the degree of solvent accessibility of the adenine residues involved in the quadruplex loops. Indeed, this property can be easily predicted from the atomic coordinates of each model by computing the solvent accessibility surface area (SASA) of each dA residues which can then be experimentally investigated by quantitative fluorescence studies using systematic single-substitutions to replace each of the adenine bases with 2-aminopurine (2-AP). Indeed, the fluorescence properties of 2-AP are extremely sensitive to local environment (conformation), making this base analog ideal for distinguishing among different quadruplex conformations [32]. This approach has been successfully

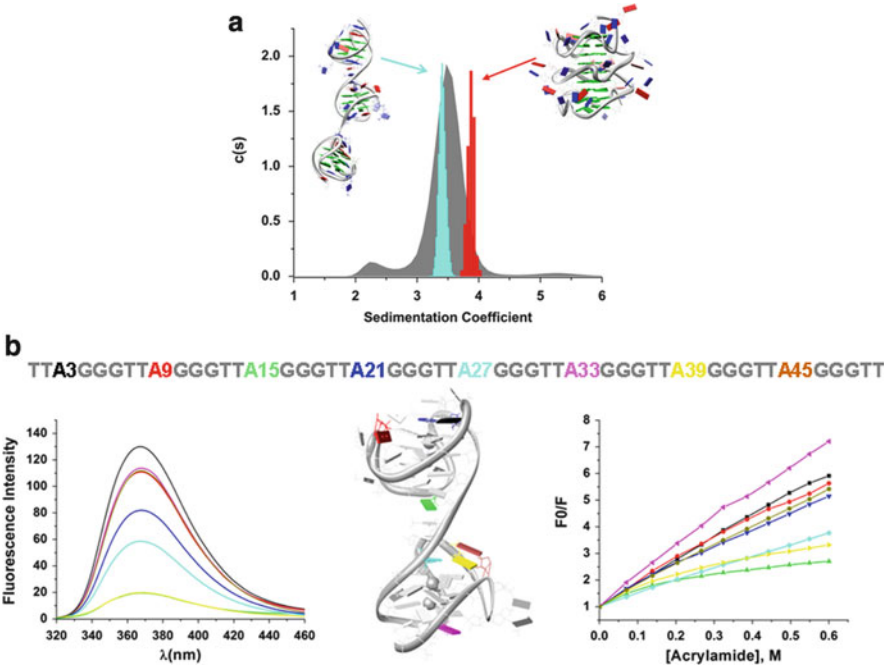


Fig. 4 Comparison between predicted properties of the higher order quadruplex models and their experimental measurement. (**a**) The predicted sedimentation coefficient distributions for the hybrid-based (*cyan*) and parallel-based (*red*) models are compared with the experimentally determined distribution (*gray*). (**b**) The results of fluorescence intensity measurements (plots on the *left*) and fluorescence quenching studies using acrylamide (plots on the *right*) are shown for oligonucleotides containing 2-aminopurine. The structure of the Hybrid-12 model is also shown (in the *middle*). The colors indicate the position of the 2-aminopurine substitution as specified in the sequence written on the *top*. Each experiment was performed in 10 mM potassium phosphate, 100 mM KCl, and 0.1 mM EDTA at pH 7.0 (from [2, 5])

employed to discriminate among the five (Hybrid-12, Hibrid-21, Hybrid-11, Hybrid-22, and all-parallel) two-quadruplex models described above (Sect. 3.2) for the eight-repeat telomeric sequence [19]. The eight-repeat sequence has eight adenine bases, each dA being in a specific loop of the two-quadruplex structure and the analysis of the five models showed that the environment of each adenine base differs depending on the folding adopted by the two-quadruplex structure. In particular, different adenine bases are involved in the junction region between the two quadruplexes depending on the model considered [19]. Both steady-state fluorescence and acrylamide quenching experiments were performed to provide a quantitative measure of the environment of the loop adenines (Fig. 4b). The trend in the fluorescence intensities and 2-AP solvent accessibilities has been shown to be consistent with the adenine environment predicted on the basis of the Hybrid-12 model, indicating that this structure can plausibly exist in solution whereas none of the other models were compatible with all the fluorescence experiments results

[19]. As shown previously, in the Hybrid-12 model the two quadruplex units are not fully independent, and are stabilized by specific interactions at the interface.

5 Energetics of Higher-Order Human Telomeric Quadruplexes Formation

The energetics of multi-quadruplex formation can provide useful information on their whole structure and on the interactions between the quadruplex subunits. Thermal stability of higher order quadruplex structures can be monitored by several spectroscopic techniques involving UV and circular dichroism (CD). In an early studies on stability of long telomeric DNA sequences, UV absorption has been employed to follow the melting behavior of the $(TTAGGG)_{n=1-12}$ series [1, 22]. These authors measured very similar melting temperatures and van't Hoff enthalpy (~190–200 kJ/mol) for the $(TTAGGG)_{n=4,8,12}$ sequences, suggesting that the 8-repeat and 12-repeat sequences are composed of quadruplex subunit similar to the quadruplex formed by the four-repeat sequence. On the other hand, similar melting temperatures are also reported for the $(TTAGGG)_n$TT series [5].

A different result is reported for the $G_3(TTAGGG)_{n=7,11,15}$ series, showing much lower T_m and ΔH° values in comparison with the corresponding $G_3(TTAGGG)_3$ single quadruplex [20]. However, it should be noted that the $G_3(TTAGGG)_3$ sequence is significantly different from the other telomeric sequences as the absence of flanking sequences at 5′-end could induce the form-3 quadruplex [14]. In contrast with the other quadruplex folding (hybrid-1, hybrid-2, and parallel), form-3 has its 5′ and 3′ end pointing in the same direction and could be unable to form linear multi-quadruplex structures. Further, the shape of the CD spectra of the $G_3(TTAGGG)_{n=7,11,15}$ series appears to be length-dependent, suggesting that the conformation of $G_3(TTAGGG)_{n=7,11,15}$ may not result from a simple addition of consecutive $G_3(TTAGGG)_3$ quadruplex units [9]. Interestingly, recent results from temperature-gradient gel electrophoresis (TGGE) experiments suggest that a significant fraction of the $G_3(TTAGGG)_7$ and $G_3(TTAGGG)_{11}$ sequences does not form the expected maximum number of quadruplex subunits [9].

Beside this, it should be noted that in the cited studies the thermodynamic parameters are derived by the van't Hoff analysis of the melting curves. The validity of this analysis is based on the two-state approximation for the melting process and it is questionable for the melting of long DNA telomeric sequences. This assumption has been experimentally tested by means of the singular value decomposition (SVD) analysis of CD spectra, demonstrating that the melting of longer telomeric sequences is clearly not a two-state process [5]. Although the van't Hoff derived parameters can provide some useful information on the average cooperative unit inside a complex structure, an accurate determination of the thermodynamics of higher order quadruplex structures should be achieved by means of model-free techniques such as DSC. This technique has been employed

to determine the thermodynamics parameters for the folding processes of the $(TTAGGG)_{n=4,8,12}$ and the $(TTAGGG)_nTT$ series [5]. Deconvolution of the thermograms by means of a statistical mechanical method unraveled that the complexity of the denaturation processes increases with the number of quadruplex units within the sequence (Fig. 5). Indeed, the DSC data can be adequately represented by two species (folded and unfolded) for the single quadruplex sequence, but there are clearly three species for the two quadruplex assembly and four species for the three quadruplex assembly. This observation suggests that each quadruplex unit in the higher-order structures is not independent and identical, but is thermodynamically unique and is influenced by its neighbors. Table 2 summarized the thermodynamic parameters for each transitions as derived by the deconvolution of the DSC profiles of the $(TTAGGG)_nTT$ series. Interestingly, once the thermodynamic parameters for the single quadruplex and the multi-quadruplex structures are known, it is possible to quantify the apparent coupling free energy [33] for the assembly of more quadruplex units. The difference between the total free energy of the folding of three contiguous quadruplexes and three times the folding free energy of a single quadruplex gives a coupling free energy ($\Delta G_{\text{Coupling}}$) of about +13 kJ/mol. The positive sign indicates an unfavorable coupling free energy in the higher order structure formation, arising from unfavorable interaction between the quadruplex units. Analysis of the energetic contributions to the total free energy change suggests that $\Delta G_{\text{Coupling}}$ contains contributions from both enthalpy and entropy components (Table 2). A similar analysis for the 8-repeat sequences shows small coupling free energies of only 2–3 kJ/mol, indicating lesser destabilizing interactions between two contiguous quadruplexes compared to 3 found in the 12-repeat sequence. It is possible that coupling free energies may become increasingly unfavorable in longer quadruplex assemblies. The ≈200 nt telomeric overhang could potentially fold into a structure with about eight contiguous quadruplex units. Larger unfavorable coupling free energies could, however, limit complete folding of the overhang, leaving single-stranded regions that might facilitate or nucleate protein binding.

Do the thermodynamic data support the presence of QQIs? Thermodynamically, the amount of all the interactions are measured by the enthalpy changes needed to unfold the structure. Theoretically, one could estimate the amount of QQI by subtracting from the total unfolding enthalpy change of the multi-quadruplex structure the sum of the enthalpy changes of the component quadruplex subunits measured separately. Although, practically, this procedure could be imprecise as the enthalpy change can be dependent on the (often unknown) quadruplex subunit conformation within the multi-quadruplex structure, a qualitative comparison of the enthalpy changes for the eight-repeat sequences with the four-repeat sequences reveals that a QQI interaction, if it exists, should be no more than 20–30 kJ/mol [5]. This value seems too small to be consistent with strong QQI as, for example, the extended stacking interactions between the G-tetrad proposed in the parallel-based model discussed previously (~70–80 kJ/mol should be expected for a stacking between two all-anti G-tetrads) [34]. However, weaker QQI as loops-mediated interactions between the consecutive quadruplexes cannot be excluded.

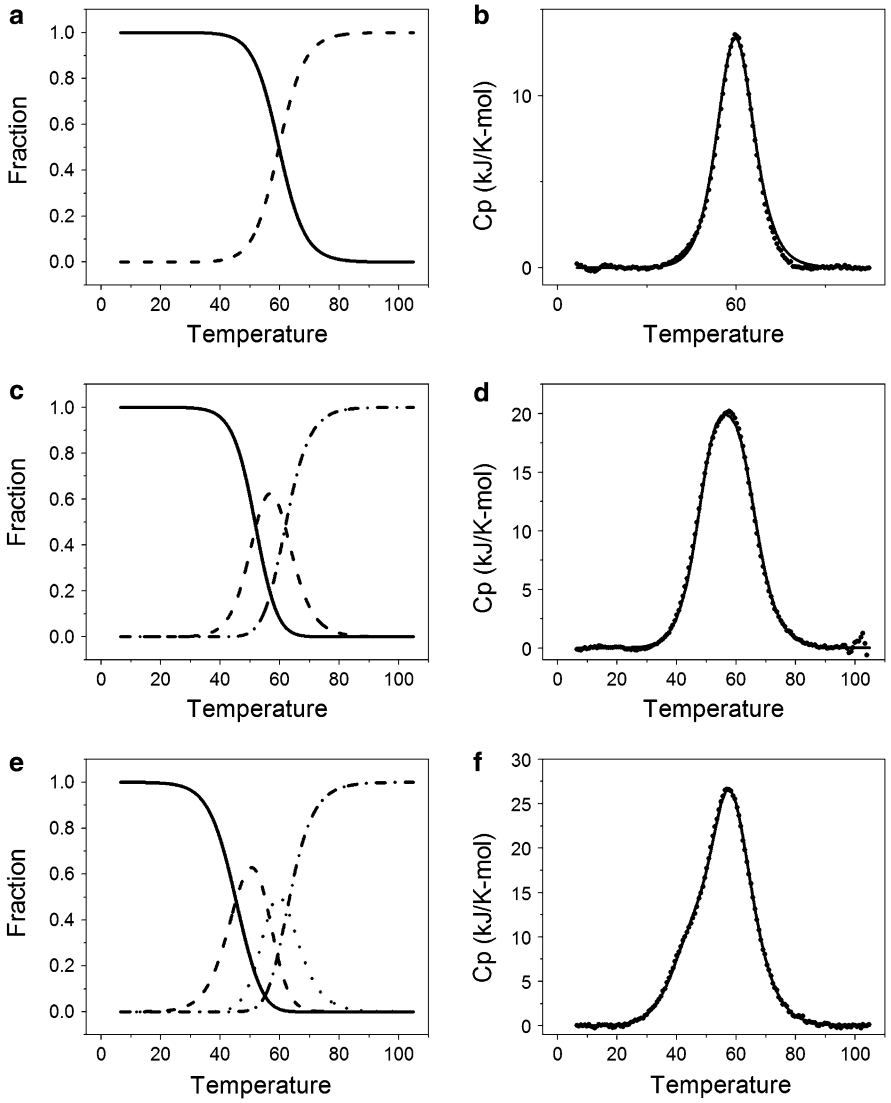


Fig. 5 Deconvolution of DSC profiles for the $(TTAGGG)_nTT$ ($n = 4,8,12$) series. Panels **a**, **c**, and **e** show the species plots for the single quadruplex, two and three quadruplexes structures, respectively. Panels **b**, **d**, and **f** show the corresponding best fits to experimental thermograms (from [5])

Interestingly, fluorescence melting experiment on an eight-repeat sequence containing 2-aminopurine shows that the loops region melts several degrees before the melting of the G-tetrad core of the quadruplexes, thus suggesting that

Table 2 Deconvolution of DSC thermograms for the $(TTAGGG)_nTT$ series[a]

Sequence	T_m (°C)	ΔH (kJ/mol)	ΔS (J/K-mol)	$\Delta G_{folding}$ (20 °C) (kJ/mol)
$(TTAGGG)_4TT$	59.8	213	639	−25.8
$(TTAGGG)_8TT$				
Transition 1	52.0	219	676	−20.9
Transition 2	61.8	222	662	−28.0
$(TTAGGG)_{12}TT$				
Transition 1	45.4	176	553	−14.0
Transition 2	56.3	221	672	−24.1
Transition 3	62.6	204	606	−25.2

[a]Data are from [5]

loops-mediated QQI could be lost before the melting of the whole multi-quadruplex structure [19].

The energetics of the quadruplex folding/unfolding processes for the eight-repeat sequence has also been evaluated in crowding/dehydrating conditions (in the presence of PEG200) by means of kinetic measurements [23]. Under these conditions, on increasing temperature, each quadruplex subunit converts from the hybrid to the parallel conformation before melting. The results suggest that folding properties of each quadruplex subunit are somewhat affected by the presence of the adjacent quadruplex.

6 Binding Properties of Higher Order Quadruplex Structures: New Prospects for Drug Design?

In the past 15 years, numerous research groups have been focused on the design of new ligands trying to optimize the interactions between these small molecules and the G-quadruplex motif [35–40]. However, in most of the G-quadruplex-ligand binding studies, the target DNA sequences are short human telomeric sequences (typically in the range 21–26 nt) able to fold into a single quadruplex structure [6, 35, 41]. Despite the increasing amount of structural information on longer DNA telomeric sequences, very few data are available on the binding properties of these sequences compared with the shorter DNA telomeric sequences. However, the existence of quadruplex–quadruplex interfaces in the longer telomeric sequences could be a predominant factor in determining the recognition properties of the telomeric DNA and the results of binding studies performed on single quadruplex may not necessarily be useful to predict the binding properties of higher order quadruplex structures. Further, the achievement of the specific recognition of the quadruplex–quadruplex interface could become an important step in the design of new quadruplex ligands able to discriminate the telomeric quadruplexes among the large number of quadruplex forming sequences existing in the human genome [42, 43].

The presence of possible binding sites involving two adjacent quadruplex units has been suggested by Neidle and coworkers [3, 44]. On the basis of crystallographic information, they built a stable model of a complex between the drug BRACO-19 and two contiguous parallel G-quadruplexes consisting of the drug sandwiched between two parallel quadruplex units. Shinohara et al., starting from a multi-quadruplex model, designed and characterized a ligand that specifically recognizes the quadruplex–quadruplex interface between two hybrid quadruplexes and that does not bind the single quadruplex unit [45]. In another study the binding of the cationic porphyrin TMPyP4 and of a triazatruxene (azatrux) to a quadruplex higher order structure was compared with the binding of the same ligands to the single quadruplex [46]. The results suggest that the binding properties of TMPyP4 and azatrux to the higher order quadruplex structure cannot be predicted on the basis of binding experiments performed on the single quadruplex. In particular, azatrux binds with a much higher affinity to the multiplex $(TTAGGG)_{12}$ than the single $(AG_3TT)_4$ quadruplex and with a stoichiometry consistent with the binding at the quadruplex–quadruplex interface. A binding study of hemin to multi-quadruplex structures further support the hypothesis of the existence of a new type of binding site (with respect to the isolated quadruplex) located at the quadruplex–quadruplex interfaces (called internal sites). Binding of hemin to the internal sites greatly enhances the hemin-DNA catalytic activity with respect to the hemin binding to an external site (not at the Q–Q interface) [17]. Binding of sanguinarine to the interface between two quadruplex units in an eight-repeat sequence has also been suggested by mass spectroscopy and Taq polymerase stop assay experiments [47].

Although still limited, taken together these preliminary studies suggest that the quadruplex subunits in the higher order structures do not behave independently, at least with respect to their binding properties. Hence, it could be more appropriate to consider the DNA telomeric structure as a higher-order structure with unique binding pockets rather than as simply a collection of single quadruplex units, during the drug-design process. However, further studies are needed to gain reliable insights into the binding behavior of the higher order quadruplex structures.

7 What About Long RNA Telomeric Sequences?

Telomeric DNA can be transcribed by DNA-dependent RNA polymerase II; the resulting RNA transcripts, ranging in size from 100 to 9,000 nucleotides, are composed of the tandem repeats of the sequence r(UUAGGG) [48–50]. All the questions/problems introduced about the long human DNA telomeric sequences arise again when considering the long human RNA sequences (TERRA). Do long RNA telomeric sequences form higher order quadruplex structures (as in DNA)? Which is the folding of the quadruplex subunits? etc. . .?

NMR, ESI–MS, and CD data suggest that RNA quadruplexes are not as polymorphic as DNA and mainly form the parallel structure both in short and long

telomeric sequences [7, 16, 51–53]. There is experimental evidence that long telomeric RNA molecules are compact and form higher-order structures based on parallel quadruplex units [54, 55]. Do the quadruplex units stack on each other or form a "beads-on-a-string" model involving non-interacting blocks? Results from recent RNase T1 cleavage experiments seem to indicate a middle way! Indeed, it was found that quadruplexes comprised of four and eight UUAGGG repeats were most resistant to RNase T1 digestion, suggesting a "beads-on-string"-like model of long human telomeric RNA, whereby each bead is made up of one or two-quadruplexes, respectively. In this model there are both interacting (inside the two-quadruplexes beads) and non-interacting quadruplexes (single quadruplex bead). However, the presence of regions with more than two stacked quadruplexes (longer beads) could not be completely excluded. A structural model involving a cation-mediated stacking of two parallel quadruplex subunits in the eight-repeat sequence r[GGG(UUAGGG)$_7$] has also been suggested by ESI–MS results [16].

Several structural models have been proposed for the RNA two-quadruplexes structure, all being based on the parallel folding and differing only in the stacking mode between the two parallel G-quadruplex blocks. There are four possible stacking modes: (1) 3′–5′, in which the stacking occurs between the 3′-end of the first quadruplex and the 5′-end of the second quadruplex; (2) 5′–3′, in which the stacking occurs between the 5′-end of the first quadruplex and the 3′-end of the second quadruplex; (3) 5′–5′, in which the stacking occurs between the 5′-end of both the quadruplex blocks; and (4) 3′–3′, in which the stacking occurs between the 3′-end of both the quadruplex blocks. In the "alternate-direction stacking" model, successive parallel quadruplexes adopt alternately the 5′–5′ and the 3′–3′ modes. Molecular modeling studies have been carried out on the 8-repeat human telomeric RNA sequence to test the feasibility of the different stacking modes between two parallel quadruplex blocks. The 5′–5′, 3′–3′, and 3′–5′ stacking modes were found to be stable throughout the MD simulations whereas the 5′–3′ stacking mode was found to be unstable, as the UUA linker was too short to span the six G-tetrads layers [55]. Continuous stacking of the loop bases and extensive loop–loop interactions, involving adenine–adenine stacking across the two blocks, were observed for the 5′–5′ stacking mode. This continuous stacking of loop bases was not observed in the simulations of other stacking modes. Experimentally, (intermolecular) 5′–5′ stacking has been observed between two parallel quadruplexes formed by the 10-mer RNA sequence r(GGGUUAGGGU) [52]. However, more stacking modes could coexist in long TERRA structures and additional experiments are needed to validate further a particular structural model.

8 Non-telomeric Higher-Order Quadruplex Structures

In the above paragraphs, computational and experimental studies on intramolecular higher-order human telomeric DNA (and RNA) quadruplex structures have been discussed. However, higher-order quadruplex structures can be found in other

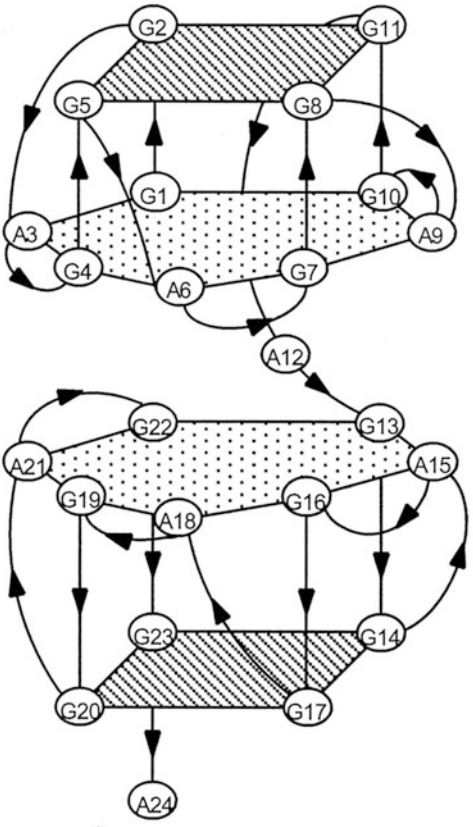


Fig. 6 Schematic representation of the higher order quadruplex structure of the $(GGA)_8$ sequence. There are extensive stacking interactions between the heptads of the quadruplex subunits (from [56])

regions of the genome. Significantly, the first demonstration of higher order packing of quadruplexes has been reported on the non-telomeric $d(GGA)_8$ sequence [56]. The GGA triplet repeat is present in eukaryotic genomes [57] and is frequently located in biologically important regions [58–62]. The four-repeat $d(GGA)_4$ sequence forms an intramolecular quadruplex composed of a G-tetrad and a G(:A):G(:A):G(A:):G heptad [63]. What has later been demonstrated is that two such G-quadruplexes can assemble, forming a two-quadruplex structure; this is, so far, the only NMR structure of a higher order quadruplex structure. The intramolecular packing of the two quadruplexes in $d(GGA)_8$ is achieved through the stacking interaction between the heptads of each quadruplex (Fig. 6). As a result, the two quadruplexes are arranged in a tail-to-tail manner. It should be noted that head-to-tail or head-to-head arrangements should involve the energetically less favorable G-tetrad-heptad or G-tetrad-G-tetrad stacking, respectively. However, sequences with higher numbers of integral multiple of four GGA repeats (i.e., with more than two quadruplex subunits) should necessarily involve these later types of quadruplex–quadruplex junctions. Interestingly, the four G–G segments (in each subunit) are aligned parallel to each other; thus the G-core of this quadruplex

architecture is similar to the G-core of the parallel human telomeric quadruplex and the structure of the $d(GGA)_8$ sequence resembles somewhat the proposed parallel-based model for the human telomeric DNA.

A higher order G-quadruplex structure has also been proposed for the hTERT, full-length, core promoter sequence containing 12 consecutive G-tracts. Taq polymerase stop assay and dimethyl sulfate footprinting experiments revealed a unique end-to-end stacked G-quadruplex structure for this sequence. This structure consists of an all parallel G-quadruplex, linked to another, atypical G-quadruplex, formed by two pairs of consecutive G-tracts separated by a 26-base loop [64].

Another example of a non-telomeric genomic region that can form multi-quadruplex structure is the insulin-linked polymorphic region (ILPR) that contains G-quadruplex forming sequences located at −363 bp upstream of the Insulin coding sequences [65–67]. It is formed by tandem repeats of the most prevalent sequence ACAGGGGTGTGGGG. It has been shown that the sequence containing two ILPR repeats ($ILPR_{n=2}$) forms an intramolecular quadruplex [66]. Although there have been no atomic-level structural studies, DMS footprinting, CD and native gel electrophoresis collectively demonstrated that the sequence containing four ILPR repeats ($ILPR_{n=4}$) forms an intramolecular structure with two G-quadruplexes [68]. Further, DMS footprinting experiments on $ILPR_{n=4}$ identified specific guanines (close to the quadruplex-quadruplex interface) protected only in the presence of the neighboring G-quadruplex, strongly suggesting some kind of QQI. TGGE results and mechanical pulling of $ILPR_{n=4}$ at the single molecule level are also consistent with the presence of QQI, showing that the two quadruplexes are unfolded cooperatively [9, 68].

In contrast with most of the findings on human telomeric DNA, CD of $ILPR_{n=4}$ is different from CD of $ILPR_{n=2}$, demonstrating that the quadruplex conformation in $ILPR_{n=4}$ is different from the quadruplex formed by $ILPR_{n=2}$, perhaps due to the presence of QQI. This result demonstrates that the $ILPR_{n=4}$ structure is far from being described as a simple addition of two $ILPR_{n=2}$ quadruplex units and more detailed structural studies are needed to get more insight into $ILPR_{n=4}$ higher order structure.

9 Concluding Remarks

Intramolecular higher-order quadruplex structures represent a challenge for biophysical and structural studies but at the same time represent a promising target for drug design that demands detailed study. Our knowledge of these higher-order structures is far from being complete and much work needs to be done. However, it is possible to delineate the following tentative conclusions from the available data:

1. Longer DNA and RNA telomeric sequences form intramolecular structures formed by single quadruplex as building blocks.

2. Most of the available data are consistent with higher order structures involving, as building blocks, hybrid-like quadruplexes (both hybrid-1 and hybrid-2) for telomeric DNA and the parallel quadruplex for telomeric RNA.
3. In DNA, the interactions between the quadruplex subunits are relatively weak and most likely based on loops-mediated interactions whereas in long RNA telomeric sequences there are large stacking interactions involving the G-tetrads of adjacent parallel quadruplexes.
4. DNA higher order quadruplex structures (and most likely RNA) have binding properties different from those of the single quadruplex subunit. Particularly, the quadruplex–quadruplex junction provides a new specific binding site.

Regarding this last point, one should emphasize that different higher order structures and quadruplex–quadruplex interfaces could be simultaneously present along the human telomeric overhang. The dynamic equilibrium among these conformations could have a key role in modulating protein recognition and in controlling the biology of the telomeres. This hypothesis awaits experimental verification. Together, the results presented in this chapter represent a good starting point for future studies on quadruplex higher-order structures aimed at a deeper understanding of this fascinating part of the quadruplex world!

Acknowledgments I thank Dr. Iolanda Fotticchia for her help in the preparation of Fig. 1. This work was supported by the "Associazione Italiana per la Ricerca sul Cancro", A.I.R.C. MFAG grant (project no. 6255).

References

1. Yu HQ, Miyoshi D, Sugimoto N (2006) Characterization of structure and stability of long telomeric DNA G-quadruplexes. J Am Chem Soc 128:15461–15468
2. Petraccone L, Trent JO, Chaires JB (2008) The tail of the telomere. J Am Chem Soc 130:16530–16532
3. Haider S, Parkinson GN, Neidle S (2008) Molecular dynamics and principal components analysis of human telomeric quadruplex multimers. Biophys J 95:296–311
4. Xu Y, Ishizuka T, Kurabayashi K et al (2009) Consecutive formation of G-quadruplexes in human telomeric-overhang DNA: a protective capping structure for telomere ends. Angew Chem Int Ed Engl 48:7833–7836
5. Petraccone L, Spink C, Trent JO et al (2011) Structure and stability of higher-order human telomeric quadruplexes. J Am Chem Soc 133:20951–20961
6. Dai J, Carver M, Yang D (2008) Polymorphism of human telomeric quadruplex structures. Biochimie 90:1172–1183
7. Phan AT (2009) Human telomeric G-quadruplex: structures of DNA and RNA sequences. FEBS J 277:1107–1117
8. Renciuk D, Kejnovska I, Skolakova P et al (2009) Arrangements of human telomere DNA quadruplex in physiologically relevant K+ solutions. Nucleic Acids Res 37:6625–6634
9. Bauer L, Tluckova K, Tohova P et al (2011) G-quadruplex motifs arranged in tandem occurring in telomeric repeats and the insulin-linked polymorphic region. Biochemistry 50:7484–7492

10. Parkinson GN, Lee MP, Neidle S (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. Nature 417:876–880
11. Ambrus A, Chen D, Dai J et al (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. Nucleic Acids Res 34:2723–2735
12. Dai J, Carver M, Punchihewa C et al (2007) Structure of the hybrid-2 type intramolecular human telomeric G-quadruplex in K+ solution: insights into structure polymorphism of the human telomeric sequence. Nucleic Acids Res 35:4927–4940
13. Luu KN, Phan AT, Kuryavyi V et al (2006) Structure of the human telomere in K+ solution: an intramolecular (3 + 1) G-quadruplex scaffold. J Am Chem Soc 128:9963–9970
14. Lim KW, Amrane S, Bouaziz S et al (2009) Structure of the human telomere in K+ solution: a stable basket-type G-quadruplex with only two G-tetrad layers. J Am Chem Soc 131:4301–4309
15. Wang Y, Patel DJ (1993) Solution structure of the human telomeric repeat d[$AG_3(T_2AG_3)_3$] G-tetraplex. Structure 1:263–282
16. Collie GW, Parkinson GN, Neidle S et al (2010) Electrospray mass spectrometry of telomeric RNA (TERRA) reveals the formation of stable multimeric G-quadruplex structures. J Am Chem Soc 132:9328–9334
17. Stefan L, Denat F, Monchaud D (2011) Deciphering the DNAzyme activity of multimeric quadruplexes: insights into their actual role in the telomerase activity evaluation assay. J Am Chem Soc 133:20405–20415
18. Wang H, Nora GJ, Ghodke H et al (2011) Single molecule studies of physiologically relevant telomeric tails reveal POT1 mechanism for promoting G-quadruplex unfolding. J Biol Chem 286:7479–7489
19. Petraccone L, Garbett NC, Chaires JB et al (2010) An integrated molecular dynamics (MD) and experimental study of higher order human telomeric quadruplexes. Biopolymers 93:533–548
20. Vorlickova M, Chladkova J, Kejnovska I et al (2005) Guanine tetraplex topology of human telomere DNA is governed by the number of (TTAGGG) repeats. Nucleic Acids Res 33:5851–5860
21. Pedroso IM, Duarte LF, Yanez G et al (2007) Sequence specificity of inter- and intramolecular G-quadruplex formation by human telomeric DNA. Biopolymers 87:74–84
22. Yu HQ, Miyoshi D, Sugimoto N (2006) Properties of long human telomeric DNAs under cell-mimicking conditions. Nucleic Acids Symp Ser (Oxf) 50:207–208
23. Petraccone L, Malafronte A, Amato J et al (2012) G-quadruplexes from human telomeric DNA: how many conformations in PEG containing solutions? J Phys Chem B 116:2294–2305
24. Sannohe Y, Sato K, Matsugami A et al (2009) The orientation of the ends of G-quadruplex structures investigated using end-extended oligonucleotides. Bioorg Med Chem 17:1870–1875
25. Xu L, Feng S, Zhou X (2011) Human telomeric G-quadruplexes undergo dynamic conversion in a molecular crowding environment. Chem Commun (Camb) 47:3517–3519
26. Lane AN (2012) The stability of intramolecular DNA G-quadruplexes compared with other macromolecules. Biochimie 94:277–286
27. Hansel R, Lohr F, Foldynova-Trantirkova S et al (2011) The parallel G-quadruplex structure of vertebrate telomeric repeat sequences is not the preferred folding topology under physiological conditions. Nucleic Acids Res 39:5768–5775
28. Miller MC, Buscaglia R, Chaires JB et al (2011) Hydration is a major determinant of the G-quadruplex stability and conformation of the human telomere 3′ sequence of d[AG(3)(TTAG(3))(3)]. J Am Chem Soc 132:17105–17107
29. Gray RD, Li J, Chaires JB (2009) Energetics and kinetics of a conformational switch in G-quadruplex DNA. J Phys Chem B 113:2676–2683
30. Li J, Correia JJ, Wang L et al (2005) Not so crystal clear: the structure of the human telomere G-quadruplex in solution differs from that present in a crystal. Nucleic Acids Res 33:4649–4659
31. Garcia De La Torre J, Huertas ML, Carrasco B (2000) Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. Biophys J 78:719–730

32. Gray RD, Petraccone L, Buscaglia R et al (2010) 2-Aminopurine as a probe for quadruplex loop structures. Methods Mol Biol 608:121–136
33. Weber G (1975) Energetics of ligand binding to proteins. Adv Protein Chem 29:1–83
34. Petraccone L, Erra E, Esposito V et al (2004) Stability and structure of telomeric DNA sequences forming quadruplexes containing four G-tetrads with different topological arrangements. Biochemistry 43:4877–4884
35. Neidle S (2010) Human telomeric G-quadruplex: the current status of telomeric G-quadruplexes as therapeutic targets in human cancer. FEBS J 277:1118–1125
36. De Cian A, Lacroix L, Douarre C et al (2008) Targeting telomeres and telomerase. Biochimie 90:131–155
37. Monchaud D, Granzhan A, Saettel N et al (2010) One ring to bind them all – part I: the efficiency of the macrocyclic scaffold for g-quadruplex DNA recognition. J Nucleic Acids 2010:pii525862
38. Granzhan A, Monchaud D, Saettel N et al (2010) One ring to bind them all – part II: identification of promising G-quadruplex ligands by screening of cyclophane-type macrocycles. J Nucleic Acids 2010:pii460561
39. Petraccone L, Barone G, Giancola C (2005) Quadruplex-forming oligonucleotides as tools in anticancer therapy and aptamers design: energetic aspects. Curr Med Chem Anticancer Agents 5:463–475
40. Petraccone L, Fotticchia I, Cummaro A et al (2011) The triazatruxene derivative azatrux binds to the parallel form of the human telomeric G-quadruplex under molecular crowding conditions: biophysical and molecular modeling studies. Biochimie 93:1318–1327
41. Burge S, Parkinson GN, Hazel P et al (2006) Quadruplex DNA: sequence, topology and structure. Nucleic Acids Res 34:5402–5415
42. Huppert JL, Balasubramanian S (2005) Prevalence of quadruplexes in the human genome. Nucleic Acids Res 33:2908–2916
43. Todd AK, Johnston M, Neidle S (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. Nucleic Acids Res 33:2901–2907
44. Haider SM, Neidle S (2009) A molecular model for drug binding to tandem repeats of telomeric G-quadruplexes. Biochem Soc Trans 37:583–588
45. Shinohara K, Sannohe Y, Kaieda S et al (2010) A chiral wedge molecule inhibits telomerase activity. J Am Chem Soc 132:3778–3782
46. Cummaro A, Fotticchia I, Franceschin M et al (2011) Binding properties of human telomeric quadruplex multimers: a new route for drug design. Biochimie 93:1392–1400
47. Bai LP, Hagihara M, Jiang ZH et al (2008) Ligand binding to tandem G quadruplexes from human telomeric DNA. Chembiochem 9:2583–2587
48. Horard B, Gilson E (2008) Telomeric RNA enters the game. Nat Cell Biol 10:113–115
49. Azzalin CM, Reichenbach P, Khoriauli L et al (2007) Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. Science 318:798–801
50. Schoeftner S, Blasco MA (2008) Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. Nat Cell Biol 10:228–236
51. Heddi B, Phan AT (2011) Structure of human telomeric DNA in crowded solution. J Am Chem Soc 133:9824–9833
52. Martadinata H, Phan AT (2009) Structure of propeller-type parallel-stranded RNA G-quadruplexes, formed by human telomeric RNA sequences in K+ solution. J Am Chem Soc 131:2570–2578
53. Arora A, Maiti S (2009) Differential biophysical behavior of human telomeric RNA and DNA quadruplex. J Phys Chem B 113:10515–10520
54. Randall A, Griffith JD (2009) Structure of long telomeric RNA transcripts: the G-rich RNA forms a compact repeating structure containing G-quartets. J Biol Chem 284:13980–13986
55. Martadinata H, Heddi B, Lim KW et al (2011) Structure of long human telomeric RNA (TERRA): G-quadruplexes formed by four and eight UUAGGG repeats are stable building blocks. Biochemistry 50:6455–6461

56. Matsugami A, Okuizumi T, Uesugi S et al (2003) Intramolecular higher order packing of parallel quadruplexes comprising a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad of GGA triplet repeat DNA. J Biol Chem 278:28147–28153
57. Beckman JS, Weber JL (1992) Survey of human and rat microsatellites. Genomics 12:627–631
58. Derry JM, Wiedemann P, Blair P et al (1995) The mouse homolog of the Wiskott–Aldrich syndrome protein (WASP) gene is highly conserved and maps near the scurfy (sf) mutation on the X chromosome. Genomics 29:471–477
59. Hirsch MR, Gaugler L, Deagostini-Bazin H et al (1990) Identification of positive and negative regulatory elements governing cell-type-specific expression of the neural cell adhesion molecule gene. Mol Cell Biol 10:1959–1968
60. Wells RD, Collier DA, Hanvey JC et al (1998) The chemistry and biology of unusual DNA structures adopted by oligopurine.oligoprimidine sequences. FASEB J 2:2939–2949
61. Chamboredon S, Briggs J, Vial E et al (2003) v-Jun downregulates the SPARC target gene by binding to the proximal promoter indirectly through Sp1/3. Oncogene 22:4047–4061
62. Palumbo SL, Memmott RM, Uribe DJ et al (2008) A novel G-quadruplex-forming GGA repeat region in the c-myb promoter is a critical regulator of promoter activity. Nucleic Acids Res 36:1755–1769
63. Matsugami A, Ouhashi K, Kanagawa M et al (2001) An intramolecular quadruplex of (GGA)(4) triplet repeat DNA with a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad, and its dimeric interaction. J Mol Biol 313:255–269
64. Palumbo SL, Ebbinghaus SW, Hurley LH (2009) Formation of a unique end-to-end stacked pair of G-quadruplexes in the hTERT core promoter with implications for inhibition of telomerase by G-quadruplex-interactive ligands. J Am Chem Soc 131:10878–10891
65. Hammond-Kosack MC, Dobrinski B, Lurz R et al (1992) The human insulin gene linked polymorphic region exhibits an altered DNA structure. Nucleic Acids Res 20:231–236
66. Hammond-Kosack MC, Kilpatrick MW, Docherty K (1993) The human insulin gene-linked polymorphic region adopts a G-quartet structure in chromatin assembled in vitro. J Mol Endocrinol 10:121–126
67. Krontiris TG (1995) Minisatellites and human disease. Science 269:1682–1683
68. Schonhoft JD, Bajracharya R, Dhakal S et al (2009) Direct experimental evidence for quadruplex–quadruplex interaction within the human ILPR. Nucleic Acids Res 37:3310–3320

Top Curr Chem (2013) 330: 47–66
DOI: 10.1007/128_2012_332

Published online: 4 July 2012

Investigation of Quadruplex Structure Under Physiological Conditions Using In-Cell NMR

Robert Hänsel, Silvie Foldynová-Trantírková, Volker Dötsch, and Lukáš Trantírek

Abstract In this chapter we describe the application of in-cell NMR spectroscopy to the investigation of G-quadruplex structures inside living *Xenopus laevis* oocytes and in *X. laevis* egg extract. First, in-cell NMR spectroscopy of nucleic acids (NA) is introduced and applications and limitations of the approach are discussed. In the following text the application of in-cell NMR spectroscopy to investigation of G-quadruplexes are reviewed. Special emphasis is given to the discussion of the influence of the intracellular environmental factors such as low molecular weight compounds, molecular crowding, and hydration on structural behavior of G-quadruplexes. Finally, future perspectives of in-cell NMR spectroscopy for quantitative characterization of G-quadruplexes and NA are discussed.

Keywords G-quadruplex · *In vivo* · In-cell NMR · Molecular crowding · *Xenopus laevis*

Contents

1 Introduction 48
2 In-Cell NMR Spectroscopy of Nucleic Acids 50
2.1 In-Cell NMR "Dactyloscopy" 52
2.2 Cellular Model: *Xenopus laevis* Oocyte/Egg 53
2.3 Application Limits 53

R. Hänsel and V. Dötsch (✉)
Institute of Biophysical Chemistry and Center for Biomolecular Magnetic Resonance, Goethe University, Frankfurt/Main, Germany
e-mail: vdoetsch@em.uni-frankfurt.de

S. Foldynová-Trantírková
Department of Medical Oncology, University Medical Center Utrecht, Utrecht, The Netherlands

L. Trantírek
Department of Chemistry, Utrecht University, Utrecht, The Netherlands
e-mail: l.trantirek@uu.nl

3 G-Quadruplexes Under Physiological Conditions 55
3.1 Influence of Low Molecular Weight Compounds 56
3.2 Effect of Native Molecular Crowding and Hydration on G-Quadruplex Structure . 58
4 Summary and Outlook 61
References 63

1 Introduction

Repetitive blocks of guanine-rich sequences have the ability to form G-quadruplex structures. Using algorithms predicting the probability of DNA sequences to form G-quadruplex structures, it has been estimated that there are over 300,000 potential quadruplex sequences (PQS) in the human genome [1]. The analysis revealed that PQS are not randomly distributed over the genome, but that they are notably concentrated in telomeric and centromeric regions as well as in promoter regions of protein coding genes, particularly in those of proto-oncogenes [1, 2]. These findings agree with experimental observations that quadruplex forming sequences contribute to maintaining genome integrity in the case of centromeric and telomeric DNA regions, and to regulation of gene expression in the case of promoter regions of protein coding genes [3]. Formation of quadruplex structures for many of these PQS was experimentally confirmed both *in vitro* and *in vivo* and for many of them the roles of these non-canonical DNA structures in DNA function and pathology were experimentally demonstrated (reviewed in [4, 5]).

Besides the biological significance of the G-quadruplex structures, their unique nanoscale geometry, biocompatibility, biodegradability, and molecular recognition capacity make them promising candidates for the construction of novel functional nanomaterials. The G-quadruplex motifs have been exploited to construct various nanostructures, such as ordered lattices, origami, supramolecular assemblies, and even three-dimensional objects [6]. In addition, G-quadruplex structures have been introduced in a variety of biomedical applications [7] including transcriptional silencing of aberrantly over-expressed genes [8].

Due to the importance of these genomic regions for both maintaining integrity of the chromosomal DNA and disease developments, the structures of the G-quadruplexes have been the subject of intense biophysical investigations including high-resolution methods such as X-Ray Diffraction (XRD) and Nuclear Magnetic Resonance (NMR) spectroscopy. However, instead of revealing unique 3D structural motifs for the G-quadruplex forming DNA, these studies provided a puzzling picture of conformational polymorphism that appeared to be promoted not only by the DNA sequence alone but also by the experimental conditions used in the XRD or NMR investigations [9, 10].

From a biophysical point of view, a property common to G-quadruplex forming sequences that has emerged from the structural investigations is their inherent sensitivity to non-specific, physical/chemical environmental factors promoting their conformational polymorphism. *In vitro* studies have demonstrated that the stability and the folding topology of the G-quadruplexes might strongly depend on

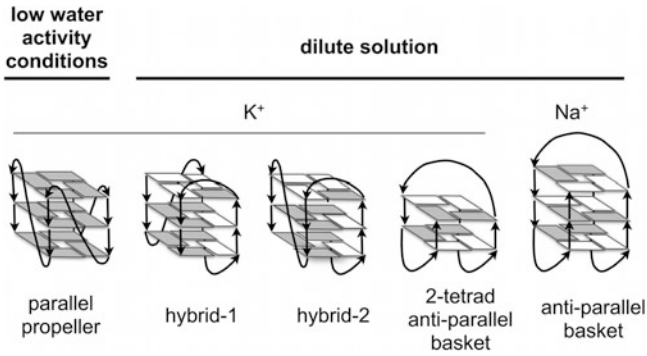


Fig. 1 Schematic representations of various human telomeric G-rich repeat structures determined under different environmental conditions. The individual drawings correspond to the parallel propeller, (3+1) parallel anti-parallel hybrid type 1, (3+1) parallel anti-parallel hybrid type 2, 2-tetrad anti-parallel basket, and anti-parallel basket G-quadruplex folds observed for d(TA $(GGGTTA)_3G_3$) in potassium solution supplemented with PEG [13], for d(TA$(GGGTTA)_3G_3$) [24], for d(TTA$(GGGTTA)_3G_3$TT) [25], and for d(($GGGTTA)_3G_3$T) in dilute potassium solutions [16], and for d($AG_3(TTAGGG)_3$) in dilute sodium solution [26], respectively. *Anti* and *syn* guanines are colored *gray* and *white*, respectively

the water activity, molecular crowding, pH, and/or nature of the counter ions [10–21]. In terms of conformational polymorphism, the G-quadruplexes formed by telomeric repeats were shown to adopt at least five distinct folding topologies, differing by virtue of strand polarities, loop orientations, and orientation of the guanine bases with respect to the sugar moiety, depending on environmental conditions and experimental setup employed for their investigation [10] (Fig. 2). Various combinations of environmental factors such as pH, molecular crowding, and cation(s) were also found to influence structural competition significantly between the G-quadruplex and the Watson–Crick duplex [22, 23]. While most of the experimental data on environmentally controlled conformational polymorphism and stability of G-quadruplexes have been acquired for telomeric sequences, they have also been observed for G-rich sequences forming quadruplex structures in promoter regions of proto-oncogenes [2, 3].

Altogether, *in vitro* structural studies have indicated that the environment-promoted structural polymorphisms and dynamic equilibrium among different folding topologies are intrinsic properties of centromeric, telomeric, and promoter DNA regions. However, at the same time, the observations that the folding topologies in these DNA regions strongly depend on subjective choice of experimental conditions and experimental setup have indicated that conventional XRD and/or NMR techniques have to be used with caution when addressing the question of which is the physiologically relevant structure of these DNA molecules. Both XRD and NMR examine the structural properties of the isolated DNA. Whereas environmental conditions in XRD studies are constrained to conditions that support monocrystal growth, the choice of NMR conditions is limited by the fact that not all relevant factors that can modulate folding topologies of these DNA structures *in vivo* are

known. Experimental observations that the folding of the G-quadruplexes depends on environmental factors have strongly suggested that quantitative characterization of their physiologically relevant structures needs to be performed in as close to native conditions as possible, preferably *in vivo*.

2 In-Cell NMR Spectroscopy of Nucleic Acids

The concept of in-cell NMR spectroscopy, which represents a crafty extension of basic principles of conventional NMR spectroscopy and allows observation of biomolecular structures in the complex cellular environment, was introduced in 2001 [27–29]. Originally, this technique was developed for selective monitoring of isotopically labeled proteins over-expressed in cultured bacterial cells. Isotopic labeling was demonstrated to be essential for discrimination of NMR signals from the molecule of the interest from signals originating from the cellular background as well as for boosting sensitivity of the in-cell NMR experiment. A principal limitation of this approach is the requirement for high-level over-expression of the protein of interest. In 2006, the original concept was extended by allowing investigation of proteins in a eukaryotic cell system, namely in *Xenopus laevis* oocytes [30, 31]. The fundamental difference between the original and new concepts is the way of introduction of the molecule to be investigated into the living cells. While in the original approach, the isotopically labeled protein is directly produced within the bacterial cells grown in isotopically labeled media, in the later approach the isotopically labeled protein is produced *ex situ* and then introduced into oocytes via microinjection. This approach provides several advantages over the original concept. First, introduction of isotopically labeled protein into isotopic label-free cellular background allows for background-free monitoring of the introduced protein. Second, this approach can be applied to proteins, which are difficult to over-express in a bacterial host. Last but not least, this approach, i.e., the introduction of exogenous molecules into these cells, can be applied to biomolecules other than proteins.

In 2009, the technique of in-cell NMR spectroscopy in injected *X. laevis* oocytes was adapted to investigate nucleic acids (NA) [32].

In contrast to the application to proteins, a specific problem of in-cell NMR of NA has been the degradation of NA inside living cells [32]. The NA, both DNA and RNA, when introduced into the cellular environment are exposed to nucleases. The nucleolytic activity results in relatively fast decay of the NA fragment under investigation. Although "life-span" of NA *in vivo* critically depends on their primary sequence and folding topology, the times accessible for DNA investigations under *in vivo* condition are typically less than 6 h [32]. Of note, the DNA forming G-quadruplex structures have represented an exception from the "rule" due to their remarkable resistance again nucleases [32]. In favorable cases, the stability of G-quadruplexes *in vivo* allows their investigation for more than 19 h under *in vivo* conditions.

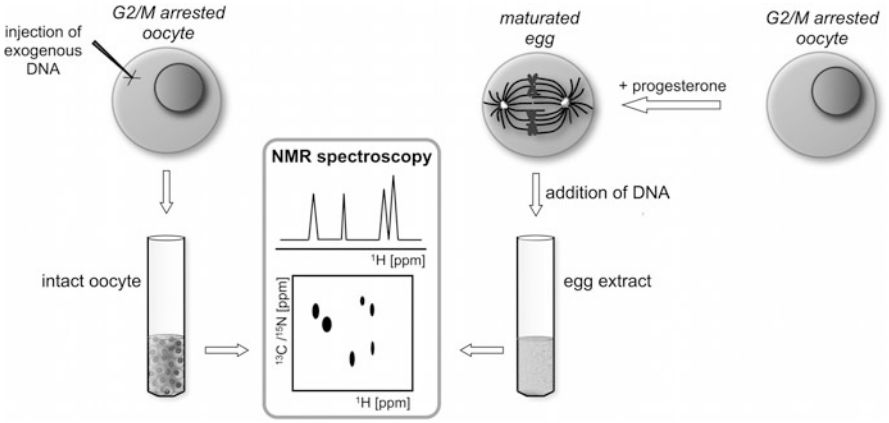


Fig. 2 Schematic representation of an in-cell NMR experiment. Cell injections with pulled glass capillaries enable the quantitative deposition of nucleic acid fragments into the intracellular environment of *X. laevis* oocytes. The in-cell NMR sample consists approximately of 200 oocytes placed into a Shigemi NMR tube. Standard 1D ^{1}H and 2D ^{1}H-^{13}C/^{15}N heteronuclear correlation NMR experiments are typically employed for collecting spectral fingerprints of unlabeled and ^{13}C/^{15}N isotopically labeled nucleic acid samples, respectively. Alternatively, NMR investigations are performed in *X. laevis* egg extract

Next to DNA degradation, life-span of the cells and incidental leakage of the introduced DNA from the cells restricts the measurement time. This time restriction dictates the type of NMR experiments than can be used for *in vivo* NA structure investigation [32]. However, one encouraging result of the original investigation was that NA inside *X. laevis* could be monitored without isotopic labeling of NA due to the fact that the cellular background does not interfere with the spectral region of imino resonances. This finding has rendered in-cell NMR analysis of unlabeled NA samples useful, as imino resonances are valuable reporters of folding topology of NA. Considering the high costs of isotopically labeled samples of NA, the fact that isotopic labeling is not required for basic structural *in vivo* analysis makes in-cell NMR an economically viable and "routine" method for investigating NA under physiological conditions [12, 32].

The basic experimental setup of an in-cell NMR experiment is schematically shown in Fig. 2. In an in-cell NMR experiment, the isotopically non-labeled or isotopically labeled NA fragment, prepared either by enzymatic or chemical synthesis, is mechanically introduced via microinjection into the cytosol of *X. laevis* oocytes. To observe and characterize quantitatively NAs inside a cell, the intracellular concentration of the introduced DNA is typically in the 50–250 μM range. For one in-cell NMR sample roughly 200 oocytes need to be injected, either manually or via a robotic device. Injected oocytes are then transferred into the NMR tube, submerged in a buffer mimicking the composition of the extra-cellular environment, and subjected to NMR investigations. Alternatively, the NMR investigations can be performed in *X. laevis* egg extract (Fig. 2).

2.1 In-Cell NMR "Dactyloscopy"

The heterogeneous nature of the cellular interior and its viscosity as well as heterogeneity of the NMR sample composed of stacked oocytes result in low resolution and S/N ratio of in-cell NMR spectra. The signals observed in in-cell NMR spectra are typically significantly broader than signals of purified samples.

In general, quantitative structure determination using NMR spectroscopy presumes the use of isotopically labeled samples coupled with exploitation of an elaborate suite of heteronuclear, multidimensional NMR experiments. While there was one reported successful protein structure determination *in vivo* using in-cell NMR spectroscopy [33], for NA, although in principle feasible, de novo structure determination *in vivo* appears impractical. The main obstacle is low *in vivo* stability of NA coupled with a prohibitively high cost for isotopically labeled samples. An additional obstacle might be toxicity. We have observed that some NA fragments were toxic to the injected cells at standard concentrations required for in-cell NMR investigations. Although lowering of the effective concentration of injected NA usually diminishes toxicity, it is inherently connected with a decreased sensitivity of the in-cell NMR experiment. As a result, the interpretation of in-cell NMR data adopted thus far has been primarily based on comparison of *in vitro* with *in vivo* spectral fingerprints rather than ab initio structure determination. The power of in-cell NMR spectroscopy thus lies not in determining structures de novo but rather in observing changes in the structures of NA due to their interactions with other cellular components.

Comparison of spectral fingerprints of NA sequences under various *in vitro* conditions promoting miscellaneous structural arrangements within studied NA fragments with *in*/*ex vivo* spectral patterns, referred to as NMR dactyloscopy, serves primarily two purposes: (1) to identify defined *in vitro* conditions, mostly a specific buffer composition, to be employed for detailed characterization of physiologically relevant structures using conventional *in vitro* NMR spectroscopy; (2) to identify the property of the intra-cellular environment that is responsible for defining of the effective conformation of the NA *in vivo*. While *in vivo* and *ex vivo* spectral patterns are recorded in living *X. laevis* oocytes and *X. laevis* egg extracts, respectively, the reference *in vitro* data are individually recorded at conditions simulating various degrees of molecular crowding, viscosity, miscellaneous pH and temperature, or different compositions of a buffer in terms of low molecular weight compounds, for example. In general this approach benefits from the high sensitivity of NMR chemical shift to both local and global structural changes of NA. Each NA folding topology is in principle marked by a unique pattern of NMR signals. As a result, an agreement between *in*/*ex vivo* and *in vitro* spectral patterns is a reliable indication of representative *in vitro* conditions under which NA fragments adopt the same folding arrangement as in a cell. Schematic representation of the interpretation process based on the comparison of NMR spectral fingerprints is outlined in Fig. 3. Practical applications of the approach are demonstrated in Sect. 3.

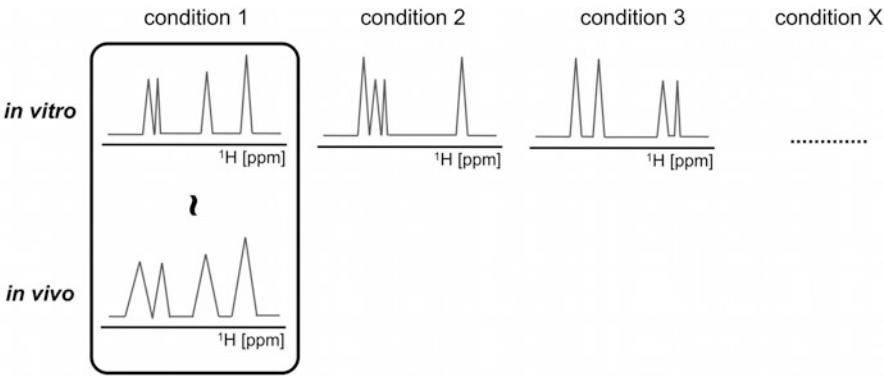


Fig. 3 Outline of interpretational process based on comparison of spectral fingerprints between *in vivo* and various *in vitro* conditions. The *in vitro* conditions are employed to generate different structural arrangements of nucleic acid fragments under study. The environmental factors such as molecular crowding, viscosity, pH, type of the counterion, or temperature are typically varied to promote distinct folding topologies of the studied sequence. Each of the unique folds is associated with the corresponding spectral fingerprint, which is subsequently compared with the spectral fingerprint recorded under physiological conditions, namely in intact *X. laevis* oocytes or *ex situ* in *X. laevis* egg extract

2.2 Cellular Model: Xenopus laevis *Oocyte/Egg*

The *X. laevis* oocyte is one of a few eukaryotic cell types that can be manipulated by microinjection on a preparative scale. *X. laevis* oocytes are widely used in many research institutions for various biochemical purposes, and protocols for their injection are well established and easily performed. Mature (stage VI) *Xenopus* oocytes have the additional advantage of being naturally synchronized and arrested in prophase at the G2/M boundary of the first meiotic division. During oocyte-to-egg maturation, a hormonal trigger activates synchronized cell-cycle progression into metaphase of meiosis II. For isolated oocytes in an *in vitro* setting, these events can be executed by the external addition of progesterone, which renders this system an important laboratory tool for the study of eukaryotic cell-cycle progression. Cellular extracts from *X. laevis* eggs are easily obtained in a largely undiluted form, and similarly recapitulate most of the biological activities of intact cells. They are frequently used as alternative cell-free systems for *ex vivo* analyses of cellular processes in many "biological" applications and are equally well suited for NMR analyses.

2.3 Application Limits

Beside the limitations mentioned above such as DNA degradation, incidental leakage, and limited cellular life-span, applications of in-cell NMR spectroscopy for NA are mainly limited by three factors: (1) low sensitivity of NMR detection,

(2) low resolution due to the non-homogenous nature of the intracellular environment and NMR sample, and (3) non-specific interactions with large components of the cellular interior. While the low sensitivity of NMR can be overcome by the introduction of high concentrations of the introduced DNA, the non-homogenous nature of the intracellular environment, NMR sample heterogeneity, and non-specific interactions generally impose limits on size of investigated systems. The restrictions and possible sources of interpretational artifacts related to these factors are discussed in detail below.

2.3.1 Intracellular Localization of Injected DNA

The very purpose of in-cell NMR is to characterize NA structure in an environment that is as close as possible to the environment of the intact cell. It follows that the setup of in-cell NMR experiments needs to reflect intra-cellular compartmentalization of the eukaryotic cell. Under native conditions, i.e., inside cells, the DNA is localized in the nucleus. It is known that low concentrations of exogenous DNA introduced into the cytoplasm of eukaryotic cells, i.e., also *X. laevis* oocytes, tend to localize spontaneously and concentrate in the nucleus [34–38]. However, as mentioned above, due to the low inherent sensitivity of NMR detection and high cellular background, the method requires large quantities (~50–250 μM) of exogenous NA to be deposited into the cellular interior [32]. In this respect, it is natural to ask whether the nucleus of a *X. laevis* oocyte is able to accommodate the high quantities of introduced DNA that are required for in-cell NMR measurements.

Our localization experiments using fluorescently labeled DNA suggested that approximately 93% of introduced DNA, at the level required for in-cell NMR experiment, is localized in the nucleus and about 7% in the cytosol. This ratio indicates that the majority of NMR signals originate from DNA localized in the nucleus.

Note that the issue of nuclear localization relates only to NMR experiments in *X. laevis* oocytes. In *X. laevis* maturated eggs, which are naturally synchronized in metaphase of meiosis II, the nuclear membrane has "ceased" to exist and the endogenous as well as introduced DNA is exposed to the same intra-cellular factors (Fig. 2). However, in terms of biological interpretations of in-cell NMR experiments in *X. laevis* oocytes and eggs, one must bear in mind that the respective spectral fingerprints relates to a specific stage of the cell-cycle, which in terms of physico-chemical properties of the intra-cellular environment can be considered as "unique" and might be different from other cell-cycle stages or from the environment in non-dividing, diploid cells [39–42].

2.3.2 Impact of Introduced DNA on Intra-cellular Environment

In their recent review on in-cell NMR spectroscopy, Ito and Selenko [43] asked the question "does being inside the cells automatically means also being native?" This question aims at the "Catch 22" of all *in vivo* techniques for NA structure

characterization, including in-cell NMR. The catch is that while these techniques were devised for biomolecular structure analysis under native condition, their implementations are always connected with various degrees of disturbance of the native environment. Without any doubt the introduction of high concentration of exogenous DNA into a living cell is inherently connected with alteration of the physico-chemical properties of the intracellular environment, which might bias the structural readout.

While deposition of large quantities of DNA into cells is an inherent attribute of in-cell NMR, the cells appear to be able to diminish adverse effects and restore homeostasis of their intracellular environment. There are two options as to how to check the integrity of manipulated cells. The first is based on a comparison of the cell viability injected with a buffer compared to cells injected with a buffer containing the DNA. The second uses the addition of progesterone to injected oocytes and checking whether the oocytes can be driven from G2/M arrested oocyte to metaphase of meiosis II, i.e., to the stage of a mature egg. As enzymatic activities required for the transition from an oocyte to a matured egg are sensitive to disturbance of environment, this simple test provides a good indication that injected cells were able to restore homeostasis of the intracellular environment.

Dramatic disturbance of the intracellular environment is typically recognized usually within 30 min from injection and is marked by increased cell mortality.

2.3.3 Size of Investigated Nucleic Acid Fragments

In general, NMR investigations of biological macromolecules are limited by the transverse relaxation time (T2). For large molecules such as NA, T2 increases proportionally with decreasing tumbling rates of the molecule in solution. The tumbling rate of the molecule decreases proportionally with increase in both the molecular mass and viscosity of the environment. It follows that viscosity of the cellular environment and size of the investigated NA fragment are among key parameters that strongly influence the sensitivity of in-cell NMR experiments since slower tumbling of a molecule causes broadening of the resonance lines which can lead to their complete disappearance.

Several lines of experimental evidence indicate that an upper limit of the intracellular viscosity for most cell types is twice that of water [44–47]. Due to the linear relationships between the viscosity, the rotational correlation time, and the molecular mass, this twofold increase in the viscosity leads to a twofold increase in the apparent molecular mass of a macromolecule, resulting in broader resonance line width.

3 G-Quadruplexes Under Physiological Conditions

Human telomeric G-rich sequences, that form monomeric intramolecular G-quadruplex structures under *in vitro* conditions, were investigated in detail using in-cell NMR spectroscopy. So far, three different telomeric repeat

constructs, forming three distinct G-quadruplex structures under *in vitro* conditions, were investigated either in intact *X. laevis* oocytes or in egg extracts [12, 32]. The outcome of these experiments revealed effects of molecular crowding and low molecular weight compounds present in the cellular environment on G-quadruplex structure. These investigations have also provided important insight into cellular hydration and put into question the physiological relevancy of crystallographic structural data as well as the data on G-quadruplex structure in solution containing polyethylene glycol (PEG) serving as a mimetic of intracellular osmolytic stress. The findings from these studies are summarized and discussed below.

3.1 Influence of Low Molecular Weight Compounds

Activities of low molecular weight compounds, particularly physiologically relevant ions such as Na^+ and K^+ and water molecules, are generally recognized as major determinants of G-quadruplex stability and structure [10, 11, 14–17, 19]. Although there is no direct experimental evidence, structural similarities between synthetic small G-quadruplex binding ligands and molecules that are natively present in the intracellular environment suggest that these might also play a role in modulating G-quadruplex structure and stability *in vivo* [48].

The human telomeric repeat sequence $d(G_3(TTAG_3)_3T)$ was the first G-quadruplex forming sequence investigated using in-cell NMR spectroscopy in living cells [32]. In potassium solution under *in vitro* conditions, this particular sequence adopts a stable 2-tetrad antiparallel basket-type structure [16]. While the *in vitro* 1D ^1H spectrum shows the expected signal pattern for the basket-type G-quadruplex [16] (Fig. 4), the imino region of the in-cell NMR spectrum recorded in injected *X. laevis* oocytes is dominated by a single, broad peak with two apparent maxima centered at 11.5 and 11.7 ppm (Fig. 4). Although the low resolution of the in-cell spectrum does not allow for a direct structural interpretation of the spectrum, its difference to the *in vitro* spectrum suggests that differences in conformation of the G-quadruplex between the *in vitro* condition and the in-cell condition might exist.

To confirm the differences and to decode their source, the injected oocytes were crushed and most of the intracellular proteins were removed from the extract by thermal denaturation. The resulting solution, so-called cleared lysate, provides improved homogeneity and a less viscous environment compared to the environment of an in-cell NMR experiment. As a result the line width of the spectrum is significantly reduced and the spectral resolution is considerably improved compared to the *in vivo* situation. However, in contrast to the situation inside cells, the cleared lysate retains only low molecular weight compounds of the intracellular environment, while both molecular crowding and confinement are discarded in the course of the lysate preparation.

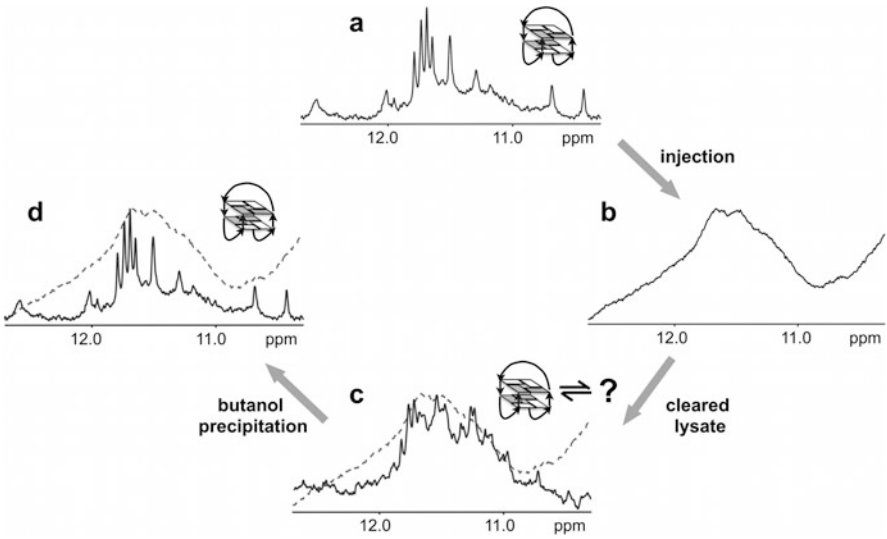


Fig. 4 Imino regions of the 1D ^{1}H 11-echo spectra of the d(G_3($TTAG_3$)$_3$T) in intraoocyte buffer (*a*), injected oocytes (*b*), cleared lysate (*c*), and buffer after butanol recovery (*d*). Note the evident difference between spectral fingerprints in intraoocyte buffer (*a*) and in the cleared lysate (*c*). While the spectrum in the buffer corresponds to single conformation, namely to the 2-tetrad anti-parallel basket, the spectrum in the lysate suggests the coexistence of multiple G-quadruplex conformations. Of note, the spectral profile in intraoocyte buffer (*a*) is quite different from *in vivo* spectral profile (marked by the *dashed line*). Figures were adapted with permission from [32]

Intriguingly, the spectral pattern of the in-cell NMR spectrum shows remarkable similarity to the 1D ^{1}H NMR spectrum recorded in cleared *X. laevis* oocyte lysate, while marked differences existed between the *in vitro* spectrum on the one hand and both in-cell and cleared lysate spectra on the other (Fig. 4). Considering the composition of the cleared lysate, the observed differences in states of the DNA between *in vitro* and cleared lysate conditions were most likely due to the G-DNA interaction with small molecules or ions. To support this conclusion, we recovered the G-DNA from the cleared lysate by butanol precipitation and dissolved it and annealed it in an *in vitro* buffer. Virtually identical patterns of the original *in vitro* spectrum and the spectrum of the G-DNA recovered from cleared lysate demonstrate that the differences between *in vitro* and cleared lysate spectra are not due to DNA degradation. In addition, the fact that the butanol precipitation is expected to "wash" out various kinds of small molecules including the ions from the G-DNA provides an additional indication that observed differences in G-quadruplex structure between *in vitro* and the cleared lysate are due to the influence of low molecular weight compounds from the cellular environment.

Of note, the similarities of spectral patterns between in-cell and cleared lysate also indirectly points to a minor role of molecular crowding in modulating G-quadruplex behavior under physiological conditions (see Sect. 3.2).

3.2 *Effect of Native Molecular Crowding and Hydration on G-Quadruplex Structure*

Large molecules, such as proteins and NA, crowd the interiors of cells. These molecules, occupying up to 40% of cellular space, can modulate structure and reactivity of other biomolecules via water exclusion effects and local confinement [49, 50]. Although the fact that the interiors of cells are crowded is nowadays generally recognized and appreciated, most conventional *in vitro* investigations on G-quadruplex structure and stability have been typically performed either under simplified, dilute conditions (CD and NMR spectroscopy) or in overcrowded, dehydrated environments in mono-crystals (X-ray crystallography). Whether any of the conditions provide a reliable representation of the native environment as it occurs inside living cells or, better, to say whether under these conditions it is possible to disclose physiologically relevant structures of G-quadruplexes has remained a major question in the field.

The role of molecular crowding in modulating the structure and stability of G-quadruplexes has been a subject of many studies. Since the techniques for *in vivo* structural analysis became available only recently, most of the studies were conducted *in vitro* in solutions supplemented mostly with synthetic organic compounds. So far, ethanol, glycerol, BSA, Ficoll, and in particular PEG have been among the most frequently used molecular crowding mimetics [12, 17, 20, 21, 49]. There were two important observations in these studies. First, different molecular crowding agents had different impacts on G-quadruplex structure and stability [12, 17, 20, 21]. This incoherence suggested that at least some of the "established" MC agents might act through non-specific physical mechanisms different from volume exclusion. Second, some of the presumed MC mimetics, namely PEG and ethanol [13, 20, 21], promoted formation of the propeller G-quadruplex structure in the presence of potassium ions, i.e., the structure that was previously observed in crystalline state, but not in dilute solutions [9]. Based on the assumption that 40% PEG200 mimics molecular crowding inside the cell, it was suggested that the propeller G-quadruplex structure might be the physiologically relevant G-quadruplex folding topology under *in vivo* conditions [13, 51]. In addition, the observation that a PEG 200 crowded environment promoted the same structural response as the crystalline state implies that the crystalline condition could resemble much closer the environment of a cell than dilute conditions as used in conventional NMR spectroscopy [51].

As recently noticed and demonstrated by Miller et al. [17] and later confirmed by our group [12], small molecules, such as PEG 200, applied at very high concentrations, act as osmolytes, decreasing water activity, rather than as mimetics of molecular crowding. These observations, have cast some doubt on the relevance of small molecules as molecular crowding agents and on the propeller G-quadruplex folding topology as the physiologically relevant structure.

In order to address the question of the physiologically relevant G-quadruplex structure(s), we recorded 1D ^{1}H NMR spectra of three different vertebra telomeric

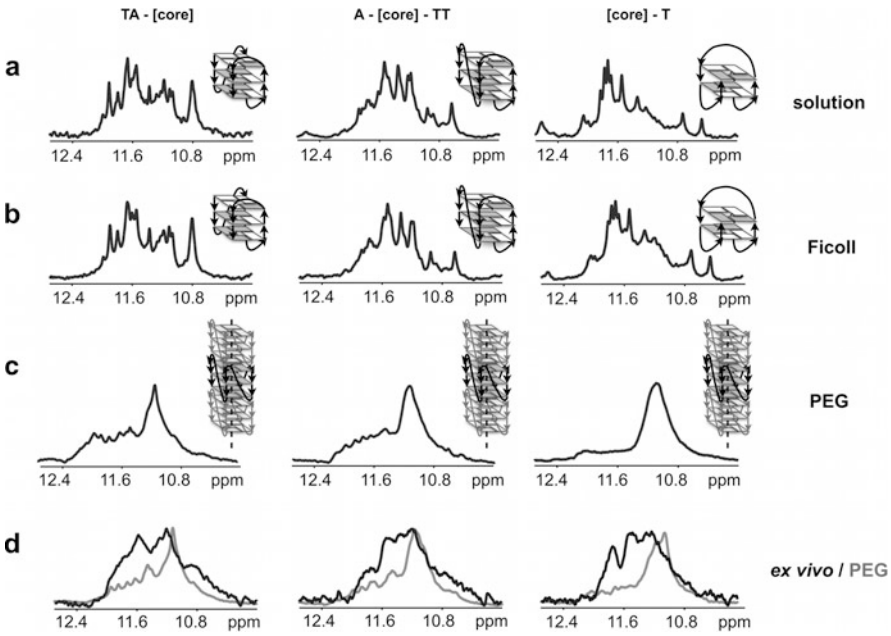


Fig. 5 Imino regions of 1D 1H 11-echo NMR spectra for d(TA(GGGTTA)$_3$G$_3$) [TA-core], d(A(GGGTTA)$_3$G$_3$TT) [A-core-TT], and [(GGGTTA)$_3$G$_3$T] [core-T] constructs in intra-oocyte buffer (**a**) in intra-oocyte buffer supplemented with 23% Ficoll 70 (**b**), in 40% PEG 200 (**c**), and in the *X. laevis* egg extract (**d**). The spectral fingerprints are accompanied by schematic representations of the assumed G-quadruplex conformations. *Anti* and *syn* guanines are colored *gray* and *white* respectively. Note the evident differences between the spectral fingerprints from *X. laevis* egg extract and those from 40% PEG 200 containing solutions. Figure adapted with permission from [12]

G-rich sequences, which adopt three distinct monomeric G-quadruplex folding topologies under dilute *in vitro* conditions, in *X. laevis* egg extract. It has been demonstrated that the egg extracts preserve many features of the intact cell and can be a good substitute for the intact cells especially as a mimetic of a native molecular crowding environment [30]. For NMR spectroscopy the measurements in the egg extract have the major advantage of improved homogeneity of the sample resulting in a reduced line width of the spectra and thus in an improved sensitivity and spectral resolution. This feature is particularly important in investigation of inherently polymorphic molecules such as G-quadruplexes. Figure 5 shows imino regions of the 1D ^{1}H NMR spectra of three different vertebra telomeric repeat sequences acquired in the *X. laevis* egg extract, in dilute potassium solution, and in the potassium solution supplemented either with 40% PEG 200 or 23% Ficoll 70. While the egg extract spectra are expected to report on physiologically relevant states of G-quadruplexes, the spectra recorded in the presence of the PEG 200 or the Ficoll 70 were both anticipated to simulate molecular crowding as it is expected inside the cells. As can be seen in Fig. 5, the signal pattern of NMR spectra for all

investigated sequences in the presence of 40% PEG 200 is considerably different from both the dilute potassium solution and the potassium solution supplemented with 23% Ficoll 70, while the NMR spectra in the diluted potassium solution and in the presence of Ficoll 70 are virtually identical. As the 40% PEG 200 and 23% Ficoll 70 exclude the same level of accessible volume, which is the primary effect of molecular crowding, the observed differences between the actions of the PEG and Ficoll indicate that at least one of them functions in a way other than a molecular crowding mimetic. Based on the definition and underlying physical chemical basis, the phenomenon of "molecular crowding" comes from the influence of large inert molecules that exclude accessible volume from the studied molecule favoring its most compact folding topology. Ficoll 70 is a neutral, highly branched, high-mass (~74 kDa), hydrophilic polysaccharide fulfilling the condition of a "large inert molecule." Although PEG has been one the most commonly used molecular crowding mimetics with NA, the PEG 200, as a rather small molecule (Mr ~ 200), does not comply with the above-mentioned condition and it might be suspected that it acts as an osmolyte, decreasing water activity and shifting conformational equilibrium towards the least hydrated form rather than as a true molecular crowding mimic. Dehydration activity of PEG 200 is also supported by the observation of similar phenotypes between PEG containing solution and other dehydrating conditions such as ethanol or inherently dehydrated crystalline environment [19, 20].

Despite the phenomenological differences, the question of how physiological relevant molecular crowding and "osmolytic stress" modulate G-quadruplex conformation and stability has remained to be addressed. The virtually identical NMR spectra between different telomeric sequences under dilute conditions and those recorded in the presence of Ficoll [12] and BSA [17] clearly indicate that excluded volume has only a minor influence on G-quadruplex state. This finding agrees with our previous observation of similar spectral fingerprints of G-quadruplex in intact *X. laevis* oocytes and cleared *X. laevis* oocytes' lysate (see above). In addition, marked differences in spectral fingerprints acquired under "dehydrating" PEG conditions and in *X. laevis* egg extract suggest that the intracellular environment is not very dehydrated in a thermodynamic sense. In other words, while the interior of a cell may be "crowded," it does not resemble either solutions of alcohols and PEG nor the crystalline state.

Although the low resolution of NMR spectra in the extract (Fig. 5) makes analysis of the G-quadruplex conformations or mixture of conformations more difficult, the fingerprint analysis clearly indicated that the propeller G-quadruplex observed under dehydration conditions, such as PEG and ethanol solutions or crystalline state, is not the physiologically prevalent folding topology.

This conclusion was recently independently supported by Tan's group, which showed that the formation of the propeller G-quadruplex in PEG containing solution, although thermodynamically favored, is also kinetically controlled [52, 53]. Xue et al. showed that the times required for propeller G-quadruplex formation notably exceed times between two rounds of cell divisions. It is expected that if G-quadruplex is to be physiologically active, its physiologically relevant

conformation needs to be formed on a time scale shorter than the time between two rounds of cell-cycle. These findings thus support the notion from in-cell NMR studies that the propeller G-quadruplex observed under dehydration conditions is not the physiologically prevalent folding topology. In addition, these findings generally emphasize the need to consider folding kinetics in addition to thermodynamic stability with respect to the biological and pharmaceutical relevance of G-quadruplex structures [52].

Altogether, our data and data from Chaires and colleagues strongly suggest that PEG, and small molecules in general, should not be used as MC mimetics and if so the interpretation of physiological relevancy of the data should be done with caution. This is particularly important for the NA community as the phenomena of "molecular crowding" and "osmolytic stress" have often been misconstrued in the literature and the confusion of these terms is usually connected with the use of ethylene glycols or alcohols, which have often been improperly referred to as molecular crowding mimetics.

4 Summary and Outlook

So far the investigation of G-quadruplex forming DNA sequences under physiological conditions either *in vivo* or *ex vivo* in intact *X. laevis* oocytes or *X. laevis* egg extracts, respectively, have indicated that conformational polymorphism of the G-quadruplex structures is primarily modulated by small molecules that are present in the intracellular environment [32]. In contrast, molecular crowding was found to have only a minor effect in tuning conformational equilibria of these inherently polymorphic molecules [12, 32]. In addition, our investigations also suggested that the parallel propeller form, which was observed under dehydrating conditions *in vitro*, was not the physiologically prevalent conformation [12]. These measurements have also suggested that although the interior of cells might be inherently crowded, it is not dehydrated. In this respect our data suggest that the results from both X-ray crystallography and NMR in the presence of dehydrating agents such as PEG or ethanol should be interpreted with extra caution with respect to their physiological relevancy.

While both *in vivo* and *ex vivo* NMR spectroscopy in intact *X. laevis* oocytes and egg extracts can provide unprecedented insight into the structural behavior of G-quadruplexes and NA under physiological conditions, the technique has several inherent limitations that need to be accounted for to avoid over-interpretation of the data. First, at the moment, the applications of in-cell NMR spectroscopy for investigation of NA are limited to *X. laevis* oocytes and eggs. In terms of physical chemical properties, the intra-cellular environment of these cells is specific for the given state of the cell cycle. The fact that the composition of intracellular milieu is known to vary with the cell cycle propagation and can be different from the environment of growing, non-dividing diploid cells has to be kept in mind when interpreting the in-cell NMR data. In addition, the data on NA behavior in one cell

type might not necessarily be the same in cells from different species. Even the cells from the same species and in the same phase of cell-cycle might display differences in terms of composition of their intracellular environment, e.g., normal vs cancer cells. Second, another limitation comes from low resolution and sensitivity of in-cell NMR spectra. At present, in-cell NMR spectroscopy is a method of choice only for small NA fragments. As recently demonstrated by Chaires and colleagues, the design/length of G-rich repeat sequences used in structural investigations critically influenced their conformational behavior [53]. Last, but not least, (in-cell) NMR primarily reports on local structure and it might be virtually "blind" to long-order changes such as environmentally induced bending of double helix [54], for example.

Undoubtedly, in-cell NMR spectroscopy of NA has a considerable application potential. However, further improvements of the technique will be necessary. One of the obvious directions for development would be extension of the approach to mammalian cells. While injection of mammalian cells appears impractical, recent development of intracellular delivery of exogenous molecules into the interior of mammalian cells, either via reversible permeabilization of the cell wall using pore-forming toxins [55] or using cargo-coupled cell-penetrating peptides [56], highlight further directions. This also appears to be the first step for the investigation of the role of the intracellular environment in the behavior of NA during individual stages of the complete cell cycle and characterization of differences between cancer and normal cells. While the problems with low resolution appear to be the inherent limitation, the use of site-specifically labeled samples is a promising solution of the problem. Recent findings that even minor changes in the compositions of commonly used buffers in conventional *in vitro* NMR measurements have notable effects on intramolecular dynamics of investigated NA [57] indicates that one of the promising applications of in-cell NMR spectroscopy of NA might be quantitative analysis of physiologically relevant NA dynamics.

As for the intrinsic "insensitivity" of in-cell NMR towards long-range structural changes, this issue cannot be properly addressed with NMR alone. On the other hand, several complementary in-cell techniques such as in-cell EPR paramagnetic electron double resonance (PELDOR) spectroscopy [58, 59] and in-cell single particle FRET techniques (Fessl et al., manuscript submitted) were recently adapted for NA. Both these techniques provide the complementary long-range information. While in-cell PELDOR is also so far limited to the *X. laevis* oocytes system, the in-cell spFRET was demonstrated for both bacterial and mammalian cells. Undoubtedly, a combination of these in-cell techniques with in-cell NMR will provide a very powerful suite of tools for solving very complex and so far unapproachable problems in cellular structural biology of NA.

Acknowledgements LT acknowledges support from a VIDI career development grant by the Netherlands Organization for Scientific Research (NOW). RH acknowledges support from the Fond der Chemischen Industrie (FCI). Support from The Center for Biomolecular Magnetic Resonance (BMRZ), the Cluster of Excellence Frankfurt (Macromolecular Complexes), and the EU (Bio-NMR) is also gratefully acknowledged.

References

1. Huppert JL, Balasubramanian S (2005) Prevalence of quadruplexes in the human genome. Nucleic Acids Res 33:2908
2. Huppert JL, Balasubramanian S (2007) G-quadruplexes in promoters throughout the human genome. Nucleic Acids Res 35:406
3. Brooks TA, Kendrick S, Hurley L (2010) Making sense of G-quadruplex and i-motif functions in oncogene promoters. FEBS J 277:3459
4. Wu Y, Brosh RM Jr (2010) G-quadruplex nucleic acids and human disease. FEBS J 277:3470
5. Bacolla A, Wells RD (2009) Non-B DNA conformations as determinants of mutagenesis and human disease. Mol Carcinog 48:273
6. Heckel A, Famulok M (2008) Building objects from nucleic acids for a nanometer world. Biochimie 90:1096
7. Collie GW, Parkinson GN (2011) The application of DNA and RNA G-quadruplexes to therapeutic medicines. Chem Soc Rev 40:5867
8. Sedoris KC, Thomas SD, Clarkson CR, Muench D, Islam A, Singh R, Miller DM (2012) Genomic c-Myc quadruplex DNA selectively kills leukemia. Mol Cancer Ther 11:66
9. Li J, Correia JJ, Wang L, Trent JO, Chaires JB (2005) Not so crystal clear: the structure of the human telomere G-quadruplex in solution differs from that present in a crystal. Nucleic Acids Res 33:4649
10. Phan AT (2010) Human telomeric G-quadruplex: structures of DNA and RNA sequences. FEBS J 277:1107
11. Dai J, Carver M, Yang D (2008) Polymorphism of human telomeric quadruplex structures. Biochimie 90:1172
12. Hansel R, Lohr F, Foldynova-Trantirkova S, Bamberg E, Trantirek L, Dotsch V (2011) The parallel G-quadruplex structure of vertebrate telomeric repeat sequences is not the preferred folding topology under physiological conditions. Nucleic Acids Res 39:5768
13. Heddi B, Phan AT (2011) Structure of human telomeric DNA in crowded solution. J Am Chem Soc 133:9824
14. Inoue M, Miyoshi D, Sugimoto N (2005) Structural switch of telomere DNA by pH and monovalent cation. Nucleic Acids Symp Ser (Oxf) 243
15. Karimata H, Miyoshi D, Sugimoto N (2005) Structure and stability of DNA quadruplexes under molecular crowding conditions. Nucleic Acids Symp Ser (Oxf) 239
16. Lim KW, Amrane S, Bouaziz S, Xu W, Mu Y, Patel DJ, Luu KN, Phan AT (2009) Structure of the human telomere in K^+ solution: a stable basket-type G-quadruplex with only two G-tetrad layers. J Am Chem Soc 131:4301
17. Miller MC, Buscaglia R, Chaires JB, Lane AN, Trent JO (2010) Hydration is a major determinant of the G-quadruplex stability and conformation of the human telomere 3' sequence of d[AG(3)[TTAG(3)](3)]. J Am Chem Soc 132:17105–17107
18. Miyoshi D, Matsumura S, Li W, Sugimoto N (2003) Structural polymorphism of telomeric DNA regulated by pH and divalent cation. Nucleosides Nucleotides Nucleic Acids 22:203
19. Parkinson GN, Lee MP, Neidle S (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. Nature 417:876
20. Rencluk D, Kejnovska I, Skolakova P, Bednarova K, Motlova J, Vorlickova M (2009) Arrangements of human telomere DNA quadruplex in physiologically relevant K^+ solutions. Nucleic Acids Res 37:6625
21. Xue Y, Kan ZY, Wang Q, Yao Y, Liu J, Hao YH, Tan Z (2007) Human telomeric DNA forms parallel-stranded intramolecular G-quadruplex in K^+ solution under molecular crowding condition. J Am Chem Soc 129:11185
22. Li W, Wu P, Ohmichi T, Sugimoto N (2002) Characterization and thermodynamic properties of quadruplex/duplex competition. FEBS Lett 526:77
23. Miyoshi D, Matsumura S, Nakano S, Sugimoto N (2004) Duplex dissociation of telomere DNAs induced by molecular crowding. J Am Chem Soc 126:165

24. Luu KN, Phan AT, Kuryavyi V, Lacroix L, Patel DJ (2006) Structure of the human telomere in K^+ solution: an intramolecular (3+1) G-quadruplex scaffold. J Am Chem Soc 128:9963
25. Dai J, Carver M, Punchihewa C, Jones RA, Yang D (2007) Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K^+ solution: insights into structure polymorphism of the human telomeric sequence. Nucleic Acids Res 35:4927
26. Wang Y, Patel DJ (1993) Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. Structure 1:263
27. Serber Z, Dotsch V (2001) In-cell NMR spectroscopy. Biochemistry 40:14317
28. Serber Z, Keatinge-Clay AT, Ledwidge R, Kelly AE, Miller SM, Dotsch V (2001) High-resolution macromolecular NMR spectroscopy inside living cells. J Am Chem Soc 123:2446
29. Serber Z, Ledwidge R, Miller SM, Dotsch V (2001) Evaluation of parameters critical to observing proteins inside living *Escherichia coli* by in-cell NMR spectroscopy. J Am Chem Soc 123:8895
30. Selenko P, Serber Z, Gadea B, Ruderman J, Wagner G (2006) Quantitative NMR analysis of the protein G B1 domain in *Xenopus laevis* egg extracts and intact oocytes. Proc Natl Acad Sci USA 103:11904
31. Serber Z, Selenko P, Hansel R, Reckel S, Lohr F, Ferrell JE Jr, Wagner G, Dotsch V (2006) Investigating macromolecules inside cultured and injected cells by in-cell NMR spectroscopy. Nat Protoc 1:2701
32. Hansel R, Foldynova-Trantirkova S, Lohr F, Buck J, Bongartz E, Bamberg E, Schwalbe H, Dotsch V, Trantirek L (2009) Evaluation of parameters critical for observing nucleic acids inside living *Xenopus laevis* oocytes by in-cell NMR spectroscopy. J Am Chem Soc 131:15761
33. Sakakibara D, Sasaki A, Ikeya T, Hamatsu J, Hanashima T, Mishima M, Yoshimasu M, Hayashi N, Mikawa T, Walchli M, Smith BO, Shirakawa M, Guntert P, Ito Y (2009) Protein structure determination in living cells by in-cell NMR spectroscopy. Nature 458:102
34. Chin DJ, Green GA, Zon G, Szoka FC Jr, Straubinger RM (1990) Rapid nuclear accumulation of injected oligodeoxyribonucleotides. New Biol 2:1091
35. Clarenc JP, Lebleu B, Leonetti JP (1993) Characterization of the nuclear binding sites of oligodeoxyribonucleotides and their analogs. J Biol Chem 268:5600
36. Fisher TL, Terhorst T, Cao X, Wagner RW (1993) Intracellular disposition and metabolism of fluorescently-labeled unmodified and modified oligonucleotides microinjected into mammalian cells. Nucleic Acids Res 21:3857
37. Leonetti JP, Mechti N, Degols G, Gagnor C, Lebleu B (1991) Intracellular distribution of microinjected antisense oligonucleotides. Proc Natl Acad Sci USA 88:2702
38. Lukacs GL, Haggie P, Seksek O, Lechardeur D, Freedman N, Verkman AS (2000) Size-dependent DNA mobility in cytoplasm and nucleus. J Biol Chem 275:1625
39. Bement WM, Capco DG (1990) Transformation of the amphibian oocyte into the egg: structural and biochemical events. J Electron Microsc Tech 16:202
40. Grandin N, Charbonneau M (1991) Cycling of intracellular free calcium and intracellular pH in Xenopus embryos: a possible role in the control of the cell cycle. J Cell Sci 99(Pt 1):5
41. Grandin N, Charbonneau M (1991) Intracellular free calcium oscillates during cell division of Xenopus embryos. J Cell Biol 112:711
42. Grandin N, Charbonneau M (1991) Changes in intracellular free calcium activity in Xenopus eggs following imposed intracellular pH changes using weak acids and weak bases. Biochim Biophys Acta 1091:242
43. Ito Y, Selenko P (2010) Cellular structural biology. Curr Opin Struct Biol 20:640
44. Guigas G, Kalla C, Weiss M (2007) The degree of macromolecular crowding in the cytoplasm and nucleoplasm of mammalian cells is conserved. FEBS Lett 581:5094
45. Guigas G, Kalla C, Weiss M (2007) Probing the nanoscale viscoelasticity of intracellular fluids in living cells. Biophys J 93:316
46. Speil J, Kubitscheck U (2010) Single ovalbumin molecules exploring nucleoplasm and nucleoli of living cell nuclei. Biochim Biophys Acta 1803:396

47. Williams SP, Haggie PM, Brindle KM (1997) ^{19}F NMR measurements of the rotational mobility of proteins in vivo. Biophys J 72:490
48. Zhang JL, Fu Y, Zheng L, Li W, Li H, Sun Q, Xiao Y, Geng F (2009) Natural isoflavones regulate the quadruplex-duplex competition in human telomeric DNA. Nucleic Acids Res 37:2471
49. Miyoshi D, Sugimoto N (2008) Molecular crowding effects on structure and stability of DNA. Biochimie 90:1040
50. Elcock AH (2010) Models of macromolecular crowding effects and the need for quantitative comparisons with experiment. Curr Opin Struct Biol 20:196
51. Borman S (2011) DNA Folding in 'crowded' conditions. Chem Eng News 89:6
52. Xue Y, Liu JQ, Zheng KW, Kan ZY, Hao YH, Tan Z (2011) Kinetic and thermodynamic control of G-quadruplex folding. Angew Chem Int Ed Engl 50:8046–8050
53. Petraccone L, Spink C, Trent JO, Garbett NC, Mekmaysy CS, Giancola C, Chaires JB (2011) Structure and stability of higher-order human telomeric quadruplexes. J Am Chem Soc 133:20951
54. Wu Z, Delaglio F, Tjandra N, Zhurkin VB, Bax A (2003) Overall structure and sugar dynamics of a DNA dodecamer from homo- and heteronuclear dipolar couplings and ^{31}P chemical shift anisotropy. J Biomol NMR 26:297
55. Ogino S, Kubo S, Umemoto R, Huang S, Nishida N, Shimada I (2009) Observation of NMR signals from proteins introduced into living mammalian cells by reversible membrane permeabilization using a pore-forming toxin, streptolysin O. J Am Chem Soc 131:10834
56. Inomata K, Ohno A, Tochio H, Isogai S, Tenno T, Nakase I, Takeuchi T, Futaki S, Ito Y, Hiroaki H, Shirakawa M (2009) High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. Nature 458:106
57. Fiala R, Spackova N, Foldynova-Trantirkova S, Sponer J, Sklenar V, Trantirek L (2011) NMR cross-correlated relaxation rates reveal ion coordination sites in DNA. J Am Chem Soc 133:13790
58. Azarkh M, Okle O, Singh V, Seemann IT, Hartig JS, Dietrich DR, Drescher M (2011) Long-range distance determination in a DNA model system inside *Xenopus laevis* oocytes by in-cell spin-label EPR. Chembiochem 12:1992
59. Krstic I, Hansel R, Romainczyk O, Engels JW, Dotsch V, Prisner TF (2011) Long-range distance measurements on nucleic acids in cells by pulsed EPR spectroscopy. Angew Chem Int Ed Engl 50:5070

Top Curr Chem (2013) 330: 67–86
DOI: 10.1007/128_2012_331

Published online: 3 July 2012

Circular Dichroism of Quadruplex Structures

Antonio Randazzo, Gian Piero Spada, and Mateus Webba da Silva

Abstract Circular dichroism (CD) is a widespread technique for studying the polymorphism of G-quadruplexes. In this chapter the CD spectral features characteristic of different folding topologies of G4-DNA are analyzed in terms of the sequence of the *syn* or *anti* glycosidic bond angle (GBA) within a quadruplex stem. Depending on the GBA sequence, the chiral disposition of two stacked guanines, adjacent along a strand, is different and this leads to a predictable contribution to the overall CD spectrum. The CD spectra of a series of G-quadruplexes, chosen as prototypal of the most common strand folding, are illustrated. The validity and the prediction power of the approach is corroborated by the analysis of CD spectra of structurally modified G4-DNA either with chemically modified guanines or polarity inversion site (5′-5′ or 3′-3′) along the strands or additional nucleobases contributing to the stacking.

Keywords Circular Dichroism · G-quadruplex · DNA · RNA

A. Randazzo (✉)
Dipartimento di Chimica Farmaceutica e Tossicologica, Università degli Studi di Napoli "Federico II", via D. Montesano 49, 80131 Napoli, Italy
e-mail: antranda@unina.it

G.P. Spada
Dipartimento di Chimica Organica "A. Mangini", Alma Mater Studiorum-Università di Bologna, via San Giacomo 11, 40126 Bologna, Italy

M. Webba da Silva
School of Biomedical Sciences, University of Ulster, Cromore Road, Coleraine BT52 1SA, UK

Contents

1 Introduction 68
2 The Origin of CD in G4-DNA 69
2.1 Group I 72
2.2 Group II 75
2.3 Group III 76
3 The CD of Chemically Modified G4-DNA 77
4 The CD Spectrum of a Quadruplex Within a Higher Order Architecture 82
References 84

1 Introduction

In the last decade the number of publications dealing with DNA G-quadruplex structures has increased exponentially, and the number of papers in which Electronic Circular Dichroism (CD) spectroscopy has been used to study quadruplex structures is extremely high [1, 2]. CD has been used to study 3D structures, ligand binding and effect of cations, to monitor the kinetics of quadruplex formation and the thermal melting, the effect of chemical modifications. However, with very few exceptions (see below) [3], these investigations use an empirical interpretation of CD spectra that relies on a relationship with spectra of known G-quadruplex structures [1, 4, 5].

Chiroptical techniques, and in particular CD [6], were developed in the second half of the last century with the main aim of assigning the absolute configuration of chiral molecules [7]. For its sensitivity to stereochemical variations, CD has emerged to become an important technique for studying subtle conformational changes and supramolecular interactions [8, 9]. At the present time the outstanding sensitivity of CD is employed mainly to study biological macromolecules and their perturbation by external factors [10] and among those G-quadruplex DNAs (G4-DNAs) represent biological structures widely investigated.

G4-DNAs are widespread in genome and it seems they play a role in a number of processes, such as replication, recombination transcription and translation [11] and are also found in telomeric regions [12]. Therefore much effort is devoted to investigate G4-DNAs along two strongly interconnected main lines: first, to understand their fine structures as a function of nucleotide sequence and external conditions (pH, temperature, solvent polarity, ionic strength, etc.) and, second, their identification as targets for chemotherapy [13, 14].

G4-DNA structures are four-stranded helical DNA structures, comprising stacks of G-quartets which are the planar association of four guanines in a cyclic Hoogsteen hydrogen-bonding arrangement (Fig. 1) [15]. G4-DNAs are formed by the folding of one nucleic acid strand or by the association of two or more strands. This results in different combinations of relative strand orientations, with consequent formation of grooves of different widths, and different loop arrangements [16]. In general, quadruplexes are called “parallel” when all strands of the

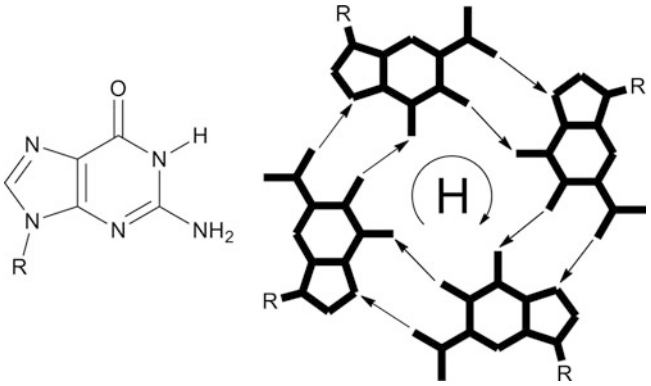


Fig. 1 The guanine moiety and the G-quartet. The G-quartet shows its "head" (H) side (from donor to acceptor H-bonding runs clockwise); the reverse side is referred as "tail" (T) (from donor to acceptor H-bonding runs counter-clockwise)

quadruplex stem have the same direction and "antiparallel" when they possess at least one strand with opposite direction.

2 The Origin of CD in G4-DNA

In general, in the case of non-coupled chromophores, the shapes of CD and UV–Vis absorption spectra are similar (except for the sign of CD bands that can be either positive or negative). On the other hand, if two (or more) strongly absorbing chromophores are chirally oriented with respect to each other, one observes an exciton spectrum characterised by the presence of two bands with opposite signs, where λ_{max} in the absorption spectrum corresponds, or nearly corresponds, to zero CD intensity [17–19]. A conservative exciton CD spectrum (equal intensity of the two bands) occurs only if the coupled chromophores possess a well isolated allowed transition; couplings with other transitions present in the same chromophore may alter the relative intensity of the two bands. An exciton CD spectrum is rather commonly observed during the formation of helical polymers or helical self-assembled aggregates. For example, CD spectra of α-helix of peptides in the absorption region of the amide chromophore have been interpreted as due to exciton coupling between $\pi-\pi^*$ transitions located in two adjacent amide groups. Similar explanations have been given for explaining the CD spectrum of B-DNA.

In the case of G4-DNA the chromophores absorbing in the UV–Vis region with $\lambda > 210$ nm are represented by guanines while contributions from the rest of the biopolymer are negligible. Guanine has in fact two well-isolated absorption bands in the 240–290 nm region which are connected to two well characterised $\pi-\pi^*$ transitions at ca. 279 and 248 nm [20, 21]; the two transitions are roughly short and long axis polarised, respectively (Fig. 2). In G4-DNAs, G-quartets are piled one on

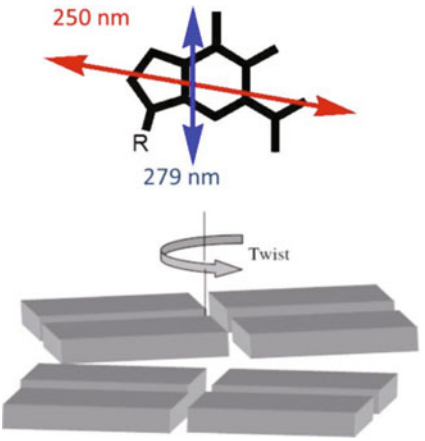


Fig. 2 Orientation of the two most relevant electric transition moments (*red* and *blue double-head arrows*) of the guanine chromophore and a sketch of the chiral arrangement of two adjacent G-quartets (each parallelepiped represents a guanine base)

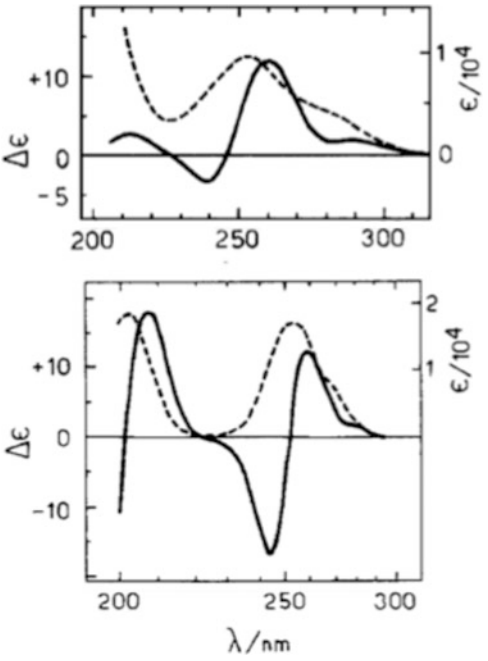


Fig. 3 *Top*: Absorption (*dashed line*) and CD (*solid line*) solution spectra of poly(G) ($\Delta\varepsilon$ is expressed in terms of nucleotide molar concentration). *Bottom*: calculated absorption (*dashed line*) and CD (*solid line*) spectra for a system composed of two G-quartets oriented as in Fig. 5, *left*. (Reproduced by permission of Società Chimica Italiana from [22])

the top of the other and rotated one with respect to the adjacent one: this rotation causes chiral exciton coupling between transition dipole moments located in near-neighbour guanines. The coupling of these moments can be calculated for plausible models and compared with experimental data.

The first non-empirical interpretation of the CD of a G4-DNA has been reported by Spada et al. [22] by modelling the spectrum of polyguanilic acid, poly(G) (Fig. 3). Poly(G) shows the typical CD spectrum of "parallel" G4-DNA dominated

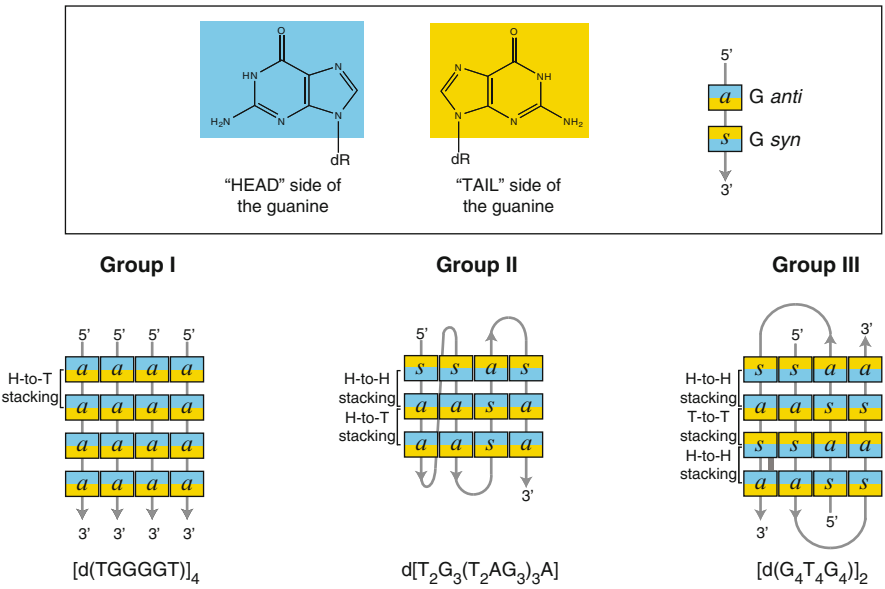


Fig. 4 A sketch of the stacking arrangement of the G-quartets observed in $[d(TG_4T)]_4$, $d[T_2G_3(T_2AG_3)_3A]$ and $[d(G_4T_4G_4)]_2$ selected as typical examples of Groups I, II and III G4-DNA, respectively (see text). Each G residue is represented by a bi-coloured rectangle and the head (H) and the tail (T) faces (as defined in Fig. 1) are *blue* and *yellow*, respectively; *s* and *a* refer to the *syn* and *anti* conformation around the glycosyl bond, respectively; the *arrow* represent the 5′-to-3′ direction of the strand. A graphic legend is reported in the box on the *top*

by a non-conservative positive exciton couplet characterised by a positive band at 260 nm and a negative one at 240 nm. This spectrum has been reproduced satisfactorily by an exciton calculation considering only near-neighbour interactions between the guanine transitions (completely neglecting the contribution from the sugar-phosphate backbone) of two stacked G-quartets [22, 23].

A G-quadruplex stem is a system of stacked guanosines that may adopt either an *anti* or a *syn* glycosidic bond angle (GBA) [24]. Accordingly, we can thus group quadruplexes [25] according to the sequence of GBA within a quadruplex stem. A quadruplex stem with strands progressing in the same direction, parallel-stranded quadruplexes, is characterised by guanosines of the same GBA (Group I) (Fig. 4). The two other possible arrangements of GBA in a quadruplex stem are characteristic of anti-parallel quadruplexes. Group II contains quadruplexes that are characterised by both sequence of guanosines of GBA of the same type (such as *anti-anti* and *syn-syn*), as well as of different type (such as *syn-anti* and *anti-syn* steps). On the other hand, Group III consists of anti-parallel quadruplexes with consecutively stacked guanosines of distinct GBA. These three groups of quadruplexes include all possible types of overlap based on GBA irrespective of the number of stacking tetrads, and are thus representative of all possible quadruplex topologies.

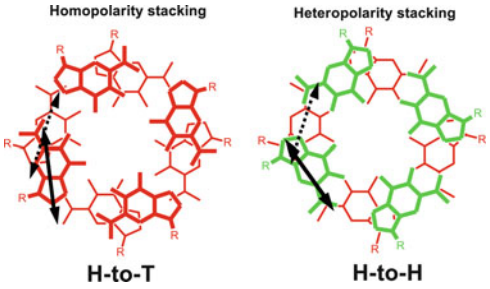


Fig. 5 Top view of the heteropolar and homopolar stacking of two G-quartets: the "head" and the "tail" sides of the G-quartets are represented in *red* and *green*, respectively (the *double-head arrows* represent the transition moments corresponding to the absorption band at ca. 250 nm)

2.1 Group I

The quadruplex $[d(TG_4T)]_4$ is known to form tetramolecular four-stranded helices based on G-quartets. The four strands run parallel in a right-handed quadruplex; the quartets are highly planar, perpendicular to the helix axis and connected to the phosphate-sugar backbone by glycosyl bonds that are exclusively *anti*, thus belonging to Group I. Due to the heteotopic nature of the two faces of the G-quartet (Fig. 1), when the G-tetrads are piled in a columnar structure, each quartet can stack onto the adjacent one through the same (head-to-head or tail-to-tail) or the opposite (head-to-tail) face leading to a heteropolar or homopolar stacking, respectively (Fig. 5). In $[d(TG_4T)]_4$ the quartets are stacked in a right-handed way and all tetrads have the same orientation, i.e. a head-to-tail (homopolar) stacking.

The CD spectrum of $[d(TG_4T)]_4$ is reported in Fig. 6. From a qualitative point of view it is not necessary to perform a detailed quanto-mechanics calculation of CD spectra. One can simply explain the positive sign of the exciton coupling corresponding to the 250-nm (i.e. positive and negative branches at 260 and 240 nm, respectively) transition of the guanine looking at the relative disposition of the two transition moments located in the closest guanines (full and dotted arrows in Fig. 5). In fact, in the framework of the chiral exciton coupling, the sign of the couplet is determined by the sign of the product $[\mathbf{r}_{ij} \bullet (\boldsymbol{\mu}_i \times \boldsymbol{\mu}_j)\, V_{ij}]$ where $\mathbf{r}_{ij}$ is the interchromophoric distance vector from i to j, $\boldsymbol{\mu}_i$ and $\boldsymbol{\mu}_j$ are the transition dipole moments on groups i and j, respectively, and V_{ij} is the interaction energy between the two dipoles [17]: knowing the position and orientation of the transition dipoles the sign of the exciton couplet can be inferred. Generally speaking, the Rotational Strength R_{0a} (a quanto-mechanical quantity that is proportional to the band area of the CD spectrum) of a specific electronic transition from ground state 0 to excited state a can be expressed as the scalar product of the electric ($\boldsymbol{\mu}_{0a}$) and magnetic ($\mathbf{m}_{0a}$) transition dipoles of the transition [7, 26]:

$$R_{0a} = \boldsymbol{\mu}_{0a} \bullet \mathbf{m}_{0a} \tag{1}$$

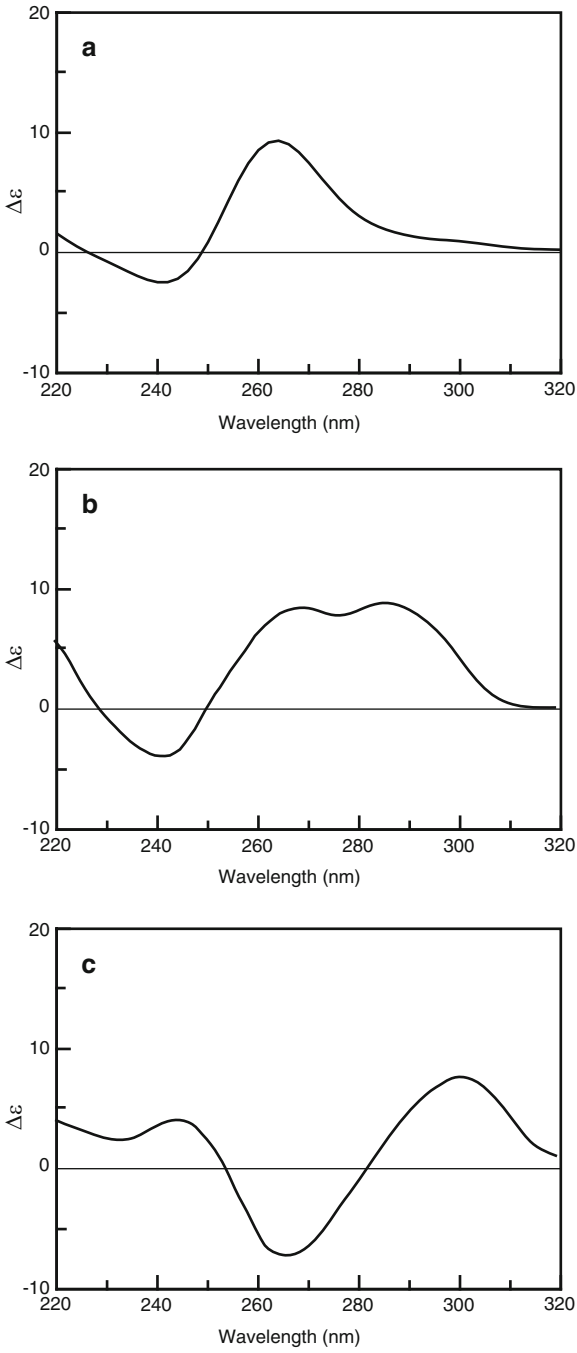


Fig. 6 CD spectra of $[d(TG_4T)]_4$ (**a**), $d[T_2G_3(T_2AG_3)_3A]$ (**b**) and $[d(G_4T_4G_4)]_2$ (**c**). CD spectra were drawn from data of [37] ($\Delta\varepsilon$ is expressed in terms of nucleotide molar concentration)

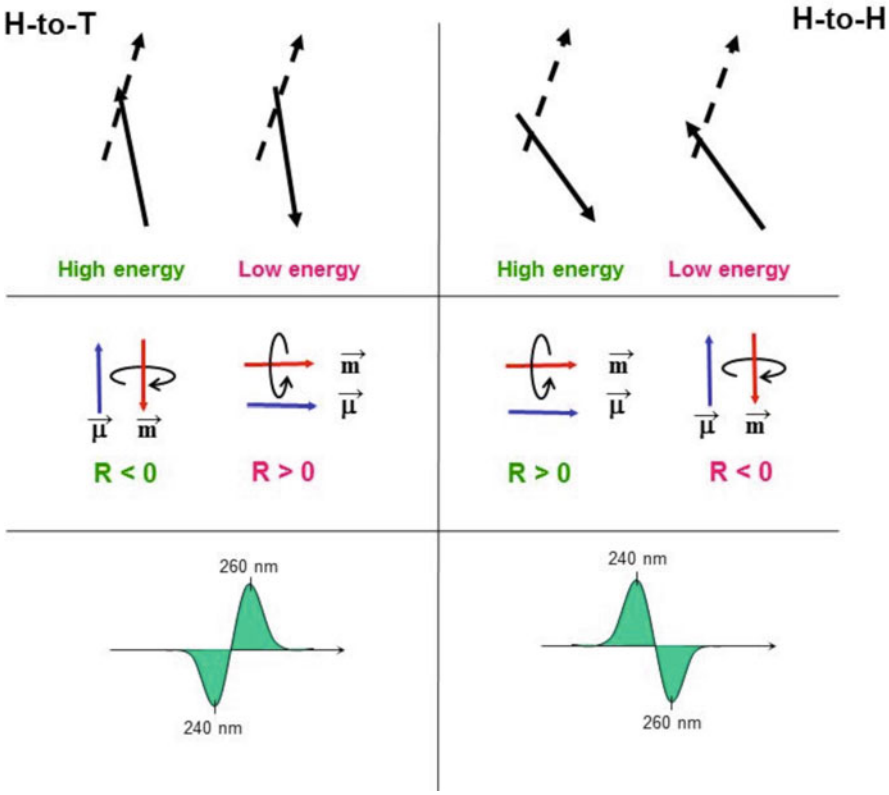


Fig. 7 A simplified model for the origin of the positive (*left*) and negative (*right*) exciton couplets for the head-to-tail (H-to-T) and head-to-head (H-to-H) G-quartet stacking, respectively. *Top*: the arrangement of two 250-nm electric transition moments (*full line*: front vector; *dashed line*: rear vector) located in two closest guanines. *Middle*: the magnetic (**m**) and electric (**μ**) moments generated by the coupling of the two guanine chromophores (more in details, in the high energy coupling of the *left* panel, the two electric transition moments – top – sum to a total electric vector (*in blue*) pointing upward middle – and generate a charge rotation with a resulting magnetic moment (in *red*) pointing downward, that is antiparallel). *Bottom*: the predicted CD spectra

Therefore, according to (1), an electronic transition must possess both electric and magnetic transition moments to be CD active.

Electric allowed transitions (as is the case for π–π* transitions in locally achiral planar aromatic chromophores) are associated with high electric dipole transition moments but negligible magnetic transition moments: the magnetic moment necessary for observing intense CD arises from the exciton coupling of two non-coplanar electric moments. The two possible coupling modes for the two electric moments (see, for example, Fig. 7) on the two chromophores are non-degenerate and give rise to magnetic moments parallel or antiparallel with respect to the total electric moment: therefore the two combinations generate positive or negative

rotational strength, respectively, the component at higher wavelength being that due to the lower energy coupling mode.

Hence, the disposition of two adjacent G-quartets in an H-to-T orientation is that reported in Fig. 7 where the electric moments (of the 250-nm transition) of a couple of near-neighbour guanines have been superposed. Applying the simplified model of the exciton coupling it emerges that this chiral arrangement (see Fig. 7, left panel) is expected to exhibit a positive exciton couplet centred at around 250 nm. This is indeed in agreement with the experimental data and the result of the QM calculations [22, 23].

2.2 Group II

The oligomer d[$T_2G_3(T_2AG_3)_3A$] belonging to Group II is characterised by three stacked quartets, generated by a folded (antiparallel) oligomer strand. The arrangement of the stacked quartets is sketched in Fig. 4: the cartoon shows that the two lower adjacent G-quartets exhibit the homopolar H-to-T stacking, while the other two adjacent G-quartets (the upper two quartets) have the heteropolar H-to-H stacking. This mixed stacking arrangement is due to the fact that the conformations of glycosyl bonds along a given G_3 tract are not all the same, being either *syn-syn-anti* or *syn-anti-anti* (in the 5′ to 3′ direction). The experimental CD spectrum of d[$T_2G_3(T_2AG_3)_3A$], see Fig. 6, is characterised by negative and positive bands at around 245 and 260 nm, respectively (very similar to Group I), and an additional positive band at ca. 290 nm. In 1998 Spada et al. [27], comparing CD spectra of different G4-DNA forming strands, proposed for the first time that the band at 290 nm "is probably due to the inversion of polarity of the quartets in the hairpin structures". Inversion of polarity (from head-to-tail to head-to-head or tail-to-tail) could cause a different arrangement of the near-neighbour transition dipole moments and consequently an oppositely signed exciton couplet corresponding to the 250-nm transition. A few years later Wen and Gray gave the same explanation [28]. They specifically proposed that a positive CD band near 260 nm arises from the stacking of quartets of the same polarity while a band at longer wavelength (290 nm) arises from the stacking of quartets with alternating polarities. When the glycosyl bonds of the guanines alternate in *syn* and *anti* conformations along each strand, the G-quartet polarity also alternates, while quadruplexes with all *anti* glycosyl bonds have non-alternating G-quartet polarity.

Only recently Gray, Fleishhauer et al. [23] calculated the CD spectrum for the quadruplex [$d(G_3T_4G_3)]_2$, that is characterised by three stacked tetrads of guanines providing one H-to-T and one heteropolar H-to-H stacking exactly like d[$T_2G_3(T_2AG_3)_3A$]. The general features of the experimental CD spectrum are all predicted well. Interestingly, the average of the spectra calculated for a pair of homo- or heteropolar stacked quartets (calculated spectra in Fig. 8) has the same spectral shape of the more refined calculated spectrum for a stack of three quartets of alternating polarities, further supporting the hypothesis that CD spectra

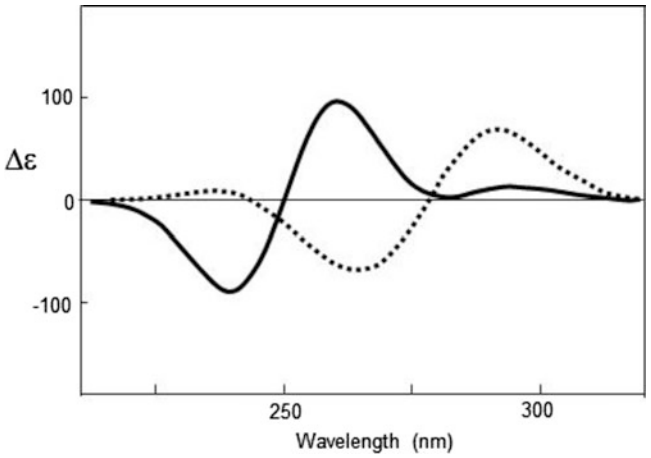


Fig. 8 Superimposed calculated CD spectra of two G-quartets stacked in the H-to-T (*solid line*) or H-to-H (*dashed line*) orientation as shown in Fig. 5. The relative orientation of the G-quartets for the calculation were extracted from the solution structure of $d(G_3T_4G_3)$ that present mixed polarities of stacked G-quartets [23]

of G-quadruplexes have shapes dominated by the stacking polarities of nearest-neighbouring G-quartets. Their results confirm the late computations by Spada et al. [26] on the homopolar stacked system and show how heteropolar stacking explains the emergence of a positive CD signal at ca. 290 nm.

2.3 Group III

As far as the bimolecular quadruplex $[d(G_4T_4G_4)]_2$ is concerned (Fig. 4), it is characterised by a CD spectrum having a positive band at 290–300 nm, whereas the other parts of the spectrum show a nearly mirror-image relationship of the bands in respect of quadruplexes belonging to Groups I and II (see Fig. 6), having a positive band at ca. 240 nm and a negative band at ca. 260 nm. This spectral difference is to be ascribed to a different stacking of the guanines. $[d(G_4T_4G_4)]_2$ is characterised by consecutively stacked guanosines of distinct glycosidic bond angles (GBA), while $d[T_2G_3(T_2AG_3)_3A]$, as mentioned above, contains sequences of guanosines of GBA of the same type (such as *anti-anti* and *syn-syn*), as well as of different type (such as *syn-anti* and *anti-syn* steps).

It should be noted that in the heteropolar H-to-H stacking of two quartets, the relative orientation of the closest dipole moments is, in fact, quite different from the H-to-T case. In Fig. 5, right, the quartets are represented in a geometry observed in many "antiparallel" G4-DNAs. Using the qualitative approach to the exciton coupling, this chiral arrangement is expected to give a negative couplet centred at around 250 nm (Fig. 7, right panel) as experimentally observed.

The simplified model of Fig. 7 could, in principle, also be used for explaining the CD signal associated with the short-axis polarized electronic transition at 279 nm; however, its lower intensity and inaccuracy in assigning its moment direction require the use of the more sophisticated QM calculations described, for example, in [22, 23].

Overall, all these observations can be summarized in the following way:

1. A band at ca. 290 nm is indicative of the presence of stacking of guanosines of different GBA
2. A positive band at ca. 240 nm coupled with a negative band at ca. 260 nm indicates the absence of stacking bases of the same GBA in the quadruplex stem
3. A negative band at ca. 240 nm coupled with a positive band at ca. 260 nm indicates the presence of stacking of guanosines of the same GBA

3 The CD of Chemically Modified G4-DNA

In the following, we will give a critical appraisal of a number of CD spectra reported in the literature that can be interpreted on the basis of the kind of stacking interactions between adjacent G-tetrads constituting the G-quadruplexes, irrespective of the relative strand orientation. In particular we will focus our attention on the CD spectra of quadruplexes containing chemical modifications that lead to the formation of unusual folding and G-tetrad stacking arrangements, for which the definition of "parallel" or "antiparallel" is not completely appropriate.

We have found the observations described in the previous section particularly useful to interpret some CD spectra of modified quadruplexes reported in the literature and not properly commented on. For example, three modified DNA quadruplexes have recently been studied, namely $[d(TG^{Br}GGT)]_4$, $[d(TGG^{Br}GT)]_4$ and $[d(TGGG^{Br}T)]_4$, where G^{Br} indicates an 8-bromoguanine residue [29]. All three quadruplexes are characterised by a fourfold symmetry with all strands parallel to each other. The set of reported data demonstrates that the replacement in turn of different guanine residues with G^{Br} in the sequence 5′-TGGGT-3′ affects the resulting quadruplex structures, leading to dissimilar CD profiles. Unlike $[d(TG^{Br}GGT)]_4$ and $[d(TGG^{Br}GT)]_4$, $[d(TGGG^{Br}T)]_4$ turned out to be not perfectly structured. Interestingly, $[d(TG^{Br}GGT)]_4$ exhibited two positive CD bands at 296 and 263 nm, and one negative band at 233 nm (see Fig. 10b), whereas the spectrum of $[d(TGG^{Br}GT)]_4$ was very similar to that of unmodified $[d(TGGGT)]_4$ and, therefore, characterised by maximum and minimum Cotton effects at 263 and 243 nm, respectively (Fig. 10a). These data suggest that G^{Br} residues in $[d(TG^{Br}GGT)]_4$ do affect the structure of the quadruplex, most probably because, due to the presence of the bulky bromine substituent in the C-8 position, these nucleobases adopt *syn* glycosyl conformations [30–32]. On the other hand, G^{Br} residue in $[d(TGG^{Br}GT)]_4$ (Fig. 9, sketch b) seems to adopt an *anti* glycosidic conformation, since the CD spectrum is basically superimposable on that of the unmodified quadruplex $[d(TGGGT)]_4$ (Fig. 9, sketch a). Fascinatingly, the

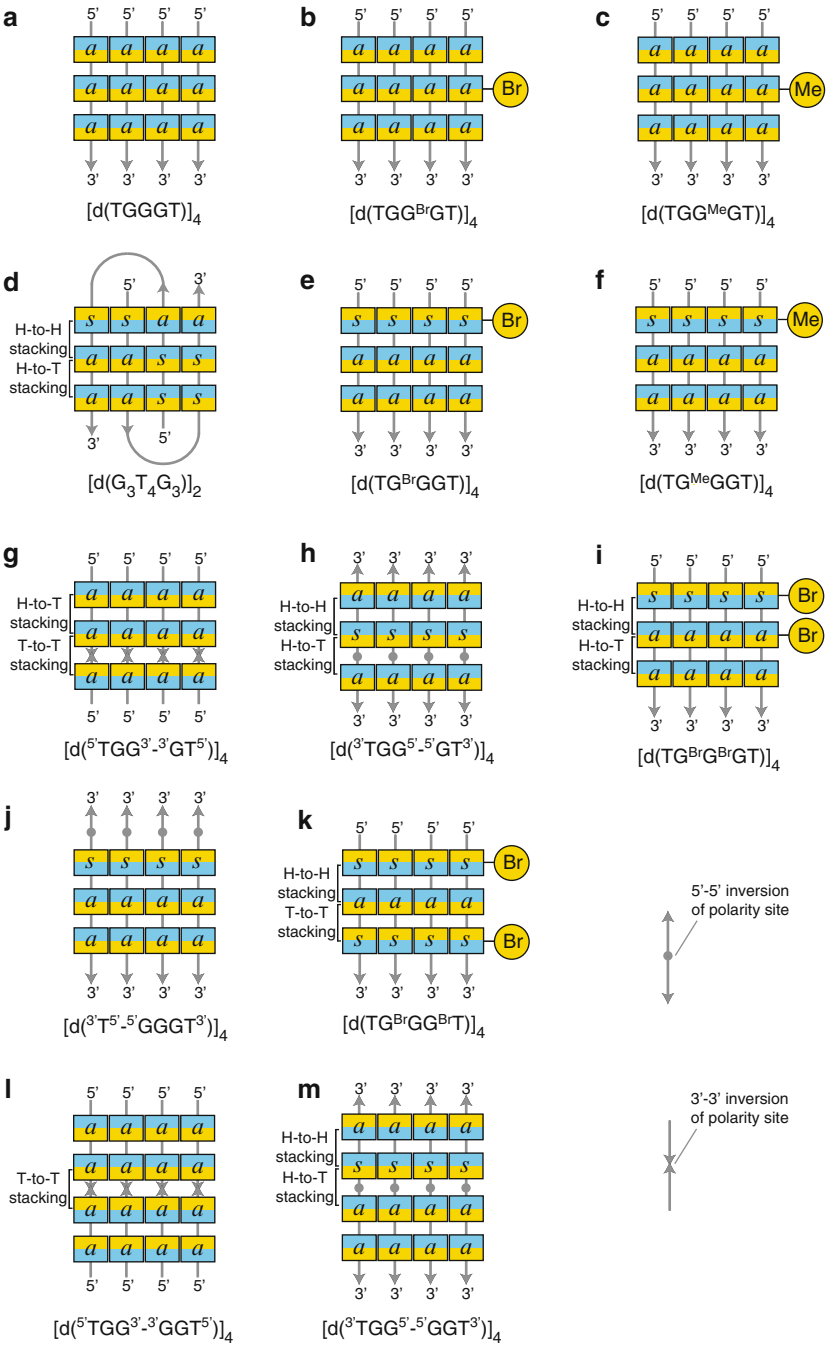


Fig. 9 A sketch of the stacking arrangement of the G-quartets observed in G-quadruplexes obtained from different oligonucleotides. For the legend, see caption to Fig. 4

"parallel" quadruplex $[d(TG^{Br}GGT)]_4$ possesses the same CD spectrum of the "antiparallel" quadruplex $[d(G_3T_4G_3)]_2$ (Group II). This apparent contradiction can be explained looking at the relative orientation of the three tetrads of the two molecules (Fig. 9, sketches d and e). Unfortunately, the NMR analysis has not been able to determine unambiguously the glycosyl bond angles of G^{Br}, due to the lack of the hydrogen in position 8 of the nucleobase. However, if we assume that the G^{Br} residues adopt a *syn* glycosidic conformation, $[d(TG^{Br}GGT)]_4$ is characterised by the two lower adjacent G-quartets exhibiting a homopolar H-to-T stacking, whereas the two upper quartets have the heteropolar H-to-H stacking (Fig. 9, sketch e), as exactly observed in $[d(G_3T_4G_3)]_2$ (Fig. 9, sketch d). Therefore, in this case, the analysis of CD data could be useful to determine the right structure of the complex. Furthermore, in a subsequent investigation, the same authors used 8-methyldeoxyguanosine residues (G^{Me}) in place of G^{Br}, taking into consideration that the methyl group has a steric size comparable to that of the bromine atom, and therefore is similarly inclined to promote the *syn* glycosyl conformation [33]. Moreover, the use of G^{Me} residues provides additional structural information based on NOE contacts involving the methyl protons in the C-8 position. As expected, the spectroscopic data of $[d(TG^{Me}GGT)]_4$ (Fig. 10b), $[d(TGG^{Me}GT)]_4$ (Fig. 10a) and $[d(TGGG^{Me}T)]_4$ turned out to be very similar to those of their analogues $[d(TG^{Br}GGT)]_4$, $[d(TGG^{Br}GT)]_4$ and $[d(TGGG^{Br}T)]_4$. However, in this case, NMR analysis was able to prove definitively that dG^{Me} adopts a *syn* glycosidic conformation in $[d(TG^{Me}GGT)]_4$ (Fig. 9, sketch f) and an *anti* glycosidic conformation in $[d(TGG^{Me}GT)]_4$ (Fig. 9, sketch c), thus inferring the interpretation of the CD spectra.

Interesting results can also be found in other investigations. Short oligonucleotides containing a site of inversion of polarity have been studied [34]. Among the six reported oligonucleotides, the sequence $^{3'}T^{5'}$-$^{5'}GGGT^{3'}$ turned out to fold into a tetramolecular quadruplex having two "parallel"-subunits joined together through the inversion of polarity sites. However, one G-tetrad adjacent to the inversion of polarity site is characterised by having all G in *syn* glycosyl conformation. Consequently, the quadruplex $[d(^{3'}T^{5'}$-$^{5'}GGGT^{3'})]_4$ (Fig. 9, sketch j) is characterised by the same number of G-tetrads having the same relative orientation as observed in $[d(G_3T_4G_3)]_2$, $[d(TG^{Br}GGT)]_4$ and $[d(TG^{Me}GGT)]_4$ (Fig. 9, sketches d–f). Thus, although $[d(^{3'}T^{5'}$-$^{5'}GGGT^{3'})]_4$ can be considered a parallel quadruplex, its CD spectrum is very similar to those of the quadruplexes indicated above belonging to Group II, showing a negative band at 233 nm and two positive bands at 265 and 295 nm (see Fig. 10d).

Another example reported in the literature deals once again with the study of modified d(TGGGT) [35]. In this case, two or three Gs have been replaced by G^{Br} residues. In particular, four modified oligonucleotides have been studied, namely d $(TG^{Br}G^{Br}GT)$, $d(TG^{Br}GG^{Br}T)$, $d(TGG^{Br}G^{Br}T)$ and $d(TG^{Br}G^{Br}G^{Br}T)$. Only the pentamers $d(TG^{Br}G^{Br}GT)$ and $d(TG^{Br}GG^{Br}T)$ were able to fold into a well-defined quadruplex structure. The CD spectrum of parallel quadruplex $[d(TG^{Br}G^{Br}GT)]_4$ (Fig. 10d) closely resembles that of the antiparallel quadruplex $[d(G_3T_4G_3)]_2$, and therefore strongly suggests that the G^{Br}-tetrad at the 5′ edge is characterised by G^{Br}

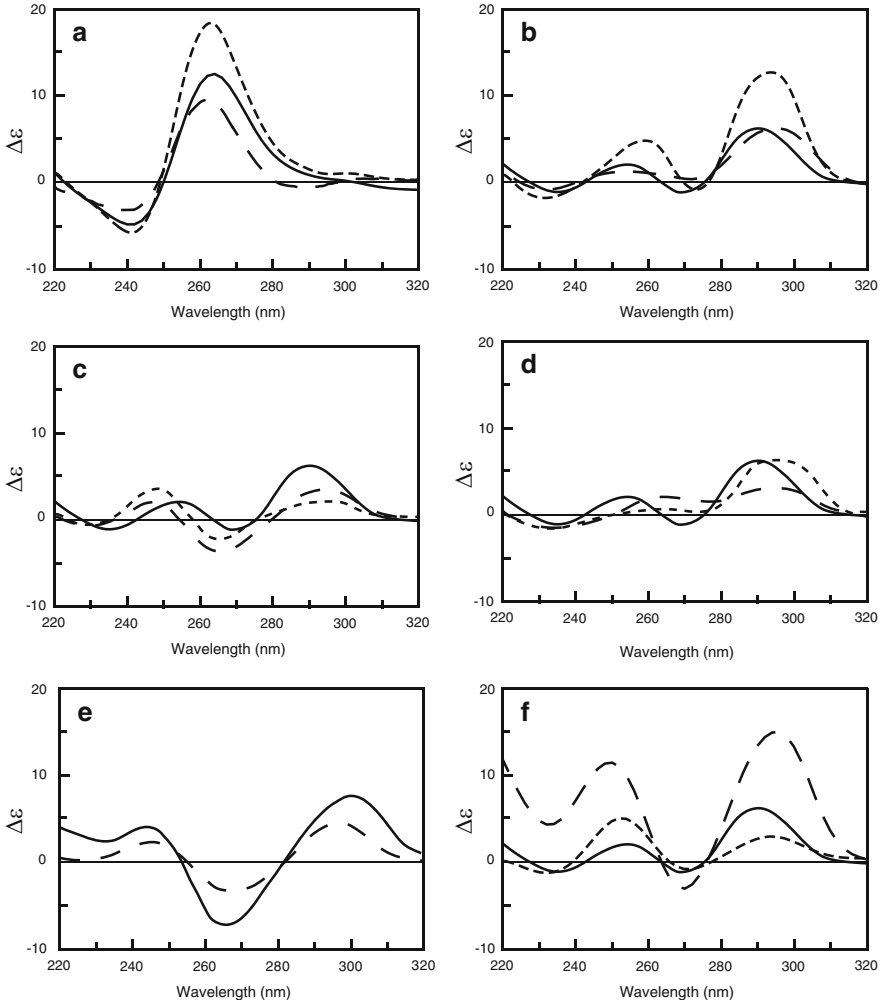


Fig. 10 CD spectra of selected G4-DNA mentioned in the text. (**a**) $[d(TGGGT)]_4$ (*solid line*), $[d(TGG^{Me}GT)]_4$ (*long-dashed line*) and $[d(TGG^{Br}GT)]_4$ (*short-dashed line*); (**b**) $[d(G_3T_4G_3)]_2$ (*solid line*), $[d(TG^{Me}GGT)]_4$ (*long-dashed line*) and $[d(TG^{Br}GGT)]_4$ (*short-dashed line*); (**c**) $[d(G_3T_4G_3)]_2$ (*solid line*), $[d(^{3'}TG^{5'}\text{-}^{5'}GGT^{3'})]_4$ (*long-dashed line*) and $[d(^{5'}TG^{3'}\text{-}^{3'}GGT^{5'})]_4$ (*short-dashed line*); (**d**) $[d(G_3T_4G_3)]_2$ (*solid line*), $[d(^{3'}T^{5'}\text{-}^{5'}GGGT^{3'})]_4$ (*long-dashed line*) and $[d(TG^{Br}G^{Br}GT)]_4$ (*short-dashed line*); (**e**) $[d(G_4T_4G_4)]_2$ (*solid line*) and $[d(TG^{Br}GG^{Br}T)]_4$ (*dashed line*); (**f**) $[d(G_3T_4G_3)]_2$ (*solid line*), $[d(^{3'}TGG^{5'}\text{-}^{5'}GGT^{3'})]_4$ (*long-dashed line*) and $[d(^{5'}TGG^{3'}\text{-}^{3'}GGT^{5'})]_4$ (*short-dashed line*). CD spectra were drawn from data of [5, 34, 36–39] ($\Delta\varepsilon$ is expressed in terms of nucleotide molar concentration)

in *syn* glycosyl conformation, whereas the G^{Br} residues of the central tetrad are supposed to adopt an *anti* glycosyl conformation (Fig. 9, sketch i). Instead, an unprecedented CD spectrum is reported for $[d(TG^{Br}GG^{Br}T)]_4$ (Fig. 10e). In particular, the presence of a positive band at 295 nm indicates that there is at least one

event of heteropolar stacking, presumably (and realistically) due to the presence of at least one *syn* G^{Br}-tetrad. It is interesting to note that, if we assume that the G^{Br}-tetrad at the 5′ edge is characterised by G^{Br} in *syn* glycosyl conformation and the one at the 3′ edge is in *anti* conformation, the overall arrangement of the tetrads should resemble that of $[d(G_3T_4G_3)]_2$ (Fig. 9, sketch d), and therefore it should give rise to a very similar CD spectrum. However, the presence of a negative band at 263 nm clearly indicates that this is not the case. Another possible arrangement should be that with a *syn* G^{Br}-tetrad at the 3′ edge and an *anti* G^{Br}-tetrad at the 5′ edge. However this arrangement would lead to one homopolar H-to-T and one heteropolar T-to-T stacking. Unfortunately no calculations have been reported so far in literature for a T-to-T stacking and, therefore, for the time being, we cannot rule out this possible arrangement. Nevertheless, the hypothesis of having both *syn* G^{Br}-tetrads at the 3′ and 5′ edges creates the classic alternation of GBA (Group III) and of the orientation of the tetrads (H-to-H and T-to-T) found in antiparallel quadruplexes like $[d(G_4T_4G_4)]_2$ (Fig. 4), whose CD spectrum closely resembles that of parallel quadruplex $[d(TG^{Br}GG^{Br}T)]_4$ (Fig. 10e).

Other examples reported in the literature of parallel quadruplexes having heteropolar (H-to-H and/or T-to-T) stacking come from other oligonucleotides having inversion of polarity sites in the middle of the G tracts. In particular, among others, the sequences $^{5'}TG^{3'}\text{-}^{3'}GGT^{5'}$, $^{3'}TG^{5'}\text{-}^{5'}GGT^{3'}$, $^{5'}TGG^{3'}\text{-}^{3'}GGT^{5'}$ and $^{3'}TGG^{5'}\text{-}^{5'}GGT^{3'}$ have been analyzed by both CD and NMR spectroscopy [34, 36]. $^{5'}TG^{3'}\text{-}^{3'}GGT^{5'}$ forms a quadruplex structure composed of two "parallel"-subunits having all the residues in *anti* glycosyl conformation. This arrangement leads to a typical H-to-T stacking with a T-to-T stacking at the inversion of polarity site (Fig. 9, sketch g). On the other hand, $[d(^{3'}TG^{5'}\text{-}^{5'}GGT^{3'})]_4$ is characterised by two "parallel"-subunits and by a *syn* G-tetrad that lead to a heteropolar H-to-H stacking (Fig. 9, sketch h) and to an H-to-T stacking at the inversion of polarity site. Interestingly, $[d(^{5'}TG^{3'}\text{-}^{3'}GGT^{5'})]_4$ and $[d(^{3'}TG^{5'}\text{-}^{5'}GGT^{3'})]_4$ are characterised by a very similar CD profile (Fig. 10c). This means that the contribution to the CD spectrum of the heteropolar stacking H-to-H is very similar to that of T-to-T stacking. The very little difference between the two spectra could be ascribed to the fact that the T-to-T stacking in $[d(^{5'}TG^{3'}\text{-}^{3'}GGT^{5'})]_4$ is right across the inversion of polarity site, where the twist angle is slightly different from that of a regular quadruplex [37]. This latter observation could also explain why their CD spectra are similar but not identical to that of $[d(G_3T_4G_3)]_2$ (Group II), which is composed of three G-tetrads having the same relative orientation.

The arrangements of $^{5'}TGG^{3'}\text{-}^{3'}GGT^{5'}$ and $^{3'}TGG^{5'}\text{-}^{5'}GGT^{3'}$ closely resemble those of $^{5'}TG^{3'}\text{-}^{3'}GGT^{5'}$ and $^{3'}TG^{5'}\text{-}^{5'}GGT^{3'}$, respectively. $^{5'}TGG^{3'}\text{-}^{3'}GGT^{5'}$ turned out to be characterised by two "parallel"-subunits with all G-tetrads having G residues in *anti* glycosyl conformation (Fig. 9, sketch l). On the other hand, $^{3'}TGG^{5'}\text{-}^{5'}GGT^{3'}$ is characterised by one G-tetrad containing all Gs in *syn* glycosyl conformation adjacent to the 5′-5′ inversion of polarity site, although it maintains a parallel organization (Fig. 9, sketch m). The interpretation of the CD spectra of $[d(^{5'}TGG^{3'}\text{-}^{3'}GGT^{5'})]_4$ and $[d(^{3'}TGG^{5'}\text{-}^{5'}GGT^{3'})]_4$ are more complicated due to the presence of inversion of polarity sites and the lack of a reference CD spectrum for

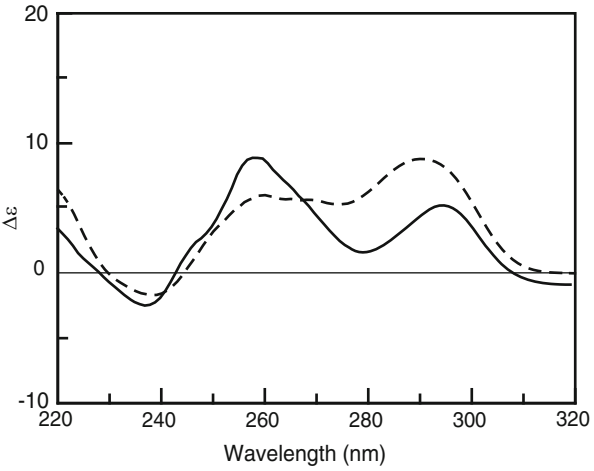


Fig. 11 CD spectra of d[$G_3(T_2AG_3)_3$ T] (*dashed line*) and d[$(TAG_3)_3TAIG_2$ (*solid line*). CD spectra were drawn from data of [38] ($\Delta\varepsilon$ is expressed in terms of nucleotide molar concentration)

these particular distributions and orientations of G-tetrads. Nevertheless, it is noteworthy that, although both quadruplexes can be considered parallel, both of them are characterised by CD spectra having a positive band at ca. 295 nm (Fig. 10f).

4 The CD Spectrum of a Quadruplex Within a Higher Order Architecture

We measured the CD spectrum of the molecule d[$G_3(T_2AG_3)_3T$] (PDB ID 2KF8) (Group III) shown in Fig. 11 [38]. Its CD spectrum has a negative band at 240 nm and a positive band at 264 nm, and not the positive band at 240 nm and negative band at 264 nm expected for a Group III topology (see Fig. 12). There is, however, a plausible explanation for this experimental result. The structure determined for this sequence in potassium solution shows a quadruplex stem composed of two stacked (G:G:G:G) tetrads. The quadruplex stem contains either *syn-anti* or *anti-syn* steps; i.e. there is no stacking of the same GBA.

However, there is extensive base-stacking of the looping residues onto the quadruplex stem. Consequently, the bases that stack the quadruplex stem form an extended higher order architecture stem. Thus, in one side of the quadruplex stem stacks a (G:G:G) triad, and on the other an (A:A:G) triad. Both triads are made up of purines with *anti* GBA. As expected, excellent intrastrand base stacking is observed between the bases of the quadruplex stem and the bases in each of the two triads. Furthermore, one of the triads is also stacked by a pyrimidine (T:T) mismatch with all bases with *anti* GBA. The other triad is stacked by a further sub-system of two stacked thymines. The extended stem of the resultant higher order architecture

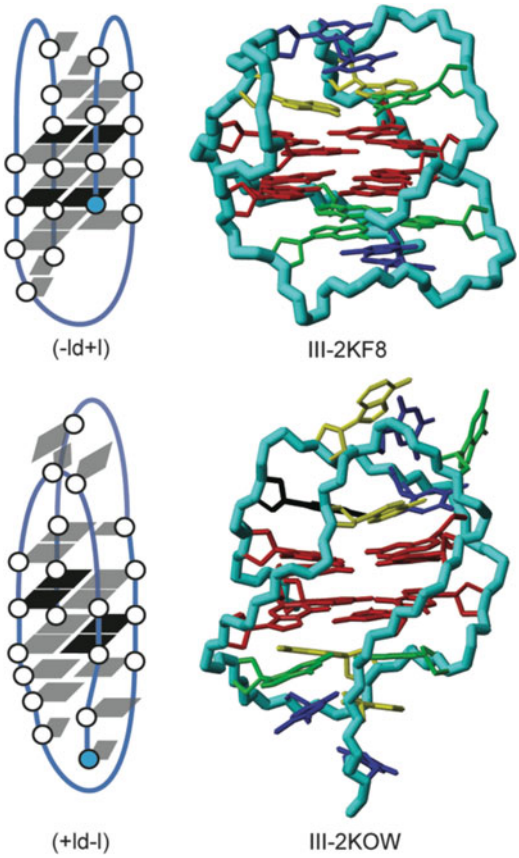


Fig. 12 *Left*: sketches of the stacking arrangement of the G-quartets observed in G-quadruplexes obtained from d[$G_3(T_2AG_3)_3$ T], *top*, and d[$(TAG_3)_3TAIG_2$], *bottom* (the 5′-end is indicated by the *shaded circle*; guanosines in *black* are *syn* and those in *grey* are *anti*; the sequence of the loops is reported in *brackets* as defined in [24, 25]). *Right*: representations for pdb IDs 2KF8 (d[$G_3(T_2AG_3)_3$ T]) and 2KOW (d[$(TAG_3)_3TAIG_2$]) (bases involved in stacking interactions are colour coded: guanines in the quadruplex stem are *red*, guanines in the triads are *green*, adenines in the triad are *yellow*, thymines in the (T:T) mismatch or stacking are *blue*)

contains three consecutively stacked guanosines with *anti* GBA – just as would be expected for Group II anti-parallel quadruplexes, and matching the observed CD signature.

A more revealing example of how stacking interactions within the stem of a higher order architecture affect the CD spectrum is described here for d[$(TAG_3)_3TAIG_2$] (PDB ID 2KOW; see Fig. 11). The sequence forms a two-stacked tetrad anti-parallel quadruplex stem with a lateral loop progressing clockwise followed by a diagonal loop, and finally another lateral loop progressing anti-clockwise (Fig. 12). The solution structure for this sequence has been determined in 90 mM potassium [39]. The CD spectrum in 10 mM potassium and 60 mM

lithium does indeed show a profile characteristic of a Group III anti-parallel quadruplex [39]: positive bands at 245 and 290 nm, and a negative band at 264 nm. However, we have measured a CD spectrum in 120 mM potassium, i.e. using experimental conditions similar to those in which the structure was determined. In this case the CD spectrum shows a negative band at 240 nm, and two positive bands at 264 and 290 nm.

There is a plausible explanation for these seemingly contradictory experimental observations. As in the case of 2KF8, the two-tetrads quadruplex stem of 2KOW in potassium is further stabilized by extensive base stacking (Fig. 12). An (A:G:G) triad stacks on one of the tetrads, and an (A:T:G) stacks on the other. The former is further stabilized by further stacking of an (A:T) base pair. All the bases in these hydrogen-bond aligned systems are in an *anti* GBA. Guanosine intrastrand stacking is thus extended from the quadruplex stem into the longer base-stacking stem of the higher order architecture. Therefore, instead of the alternating *syn-anti* or *anti-syn*, this longer stem contains *syn-anti-anti* and *anti-anti-syn* intrastrand stacking of guanosines. This is what would be expected for a Group II anti-parallel quadruplex stem with a characteristic CD profile denoting a negative band at 240 nm, and two positive bands at 264 and 290 nm. Thus, the spectrum reported for this sequence in a mixture containing 10 mM potassium and 60 mM lithium indicates that lithium destabilized the hydrogen-bond alignments stacking onto the quadruplex stem.

References

1. Paramasivan S, Rujan I, Bolton PH (2007) Circular dichroism of quadruplex DNAs: applications to structure, cation effects and ligand binding. Methods 43:324–331
2. Burge S, Parkinson GN, Hazel P, Todd AK, Neidle S (2006) Quadruplex DNA: sequence, topology and structure. Nucleic Acids Res 34:5402–5415
3. Masiero S, Trotta R, Pieraccini S, De Tito S, Perone R, Randazzo A, Spada GP (2010) A non-empirical chromophoric interpretation of CD spectra of DNA G-quadruplex structures. Org Biomol Chem 8:2683–2692
4. Kypr J, Kejnovská I, Renčiuk D, Vorlíčková M (2009) Circular dichroism and conformational polymorphism of DNA. Nucleic Acids Res 37:1713–1725
5. Jaumot J, Eritja R, Navea S, Gargallo R (2009) Classification of nucleic acids structures by means of the chemometric analysis of circular dichroism spectra. Anal Chim Acta 642:117–126
6. Berova N, Nakanishi K, Woody RW (2000) Circular dichroism: principles and applications, 2nd edn. Wiley-VCH, New York
7. Berova N, Di Bari L, Pescitelli G (2007) Application of electronic circular dichroism in configurational and conformational analysis of organic compounds. Chem Soc Rev 36:914–931
8. Gottarelli G, Lena S, Masiero S, Pieraccini S, Spada GP (2008) The use of circular dichroism spectroscopy for studying the chiral molecular self-assembly: an overview. Chirality 20:471–485
9. van Dijk L, Bobbert PA, Spano FC (2010) Extreme sensitivity of circular dichroism to long-range excitonic couplings in helical supramolecular assemblies. J Phys Chem B 114:817–825
10. Fasman GD (1996) Circular dichroism and the conformational analysis of biomolecules. Plenum, New York

11. Johnson JE, Smith JS, Kozak ML, Johnson FB (2008) In vivo veritas: using yeast to probe the biological functions of G-quadruplexes. Biochimie 90:1250–1263
12. Henderson E, Hardin CC, Walk SK, Tinoco I Jr, Blackburn EH (1987) Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine·guanine base pairs. Cell 51:899–908
13. Neidle S, Parkinson GN (2002) Telomere maintenance as a target for anticancer drug discovery. Nat Rev Drug Discov 1:383–393
14. Incles CM, Schultes CM, Neidle S (2003) Telomerase inhibitors in cancer therapy: current status and future directions. Curr Opin Investig Drugs 4:675–685
15. Neidle S, Balasubramanian S (2006) Quadruplex nucleic acids. RSC, Cambridge
16. Parkinson GN (2006) Fundamentals of quadruplex structures. In: Neidle S, Balasubramanian S (eds) Quadruplex nucleic acids. RSC, Cambridge
17. Harada H, Nakanishi K (1983) Circular dichroic spectroscopy – exciton coupling in organic stereochemistry. University Science Book, Mill Valley
18. Berova N, Nakanishi K (2000) Principles and applications of exciton chirality method. In: Berova N, Nakanishi K, Woody RW (eds) Circular dichroism – principles and applications, 2nd edn. Wiley-VCH, New York
19. Superchi S, Giorgio E, Rosini C (2004) Structural determinations by circular dichroism spectra analysis using coupled oscillator methods: an update of the applications of the DeVoe polarizability model. Chirality 16:422–451
20. Clark LB (1994) Electronic spectra of crystalline guanosine: transition moment directions of the guanine chromophore. J Am Chem Soc 116:5265–5270
21. Füelsher MP, Serrano-Andrés L, Roos BO (1997) A theoretical study of the electronic spectra of adenine and guanine. J Am Chem Soc 119:6168–6176
22. Gottarelli G, Palmieri P, Spada GP (1990) The exciton optical activity of the four-stranded helix of poly(G). Gazz Chim Ital 120:101–107
23. Gray DM, Wen JD, Gray CW, Repges R, Repges C, Raabe G, Fleischhauer J (2008) Measured and calculated CD spectra of G-quartets stacked with the same or opposite polarities. Chirality 20:431–440
24. Webba da Silva M (2007) Geometric formalism for DNA quadruplex folding. Chemistry 13:9738–9745
25. Webba da Silva M, Trajkovski M, Sannohe Y, Ma'ani Hessari N, Sugiyama H, Plavec J (2009) Design of a G-quadruplex topology through glycosidic bond angles. Angew Chem 121:9331–9334; Angew Chem Int Ed 48:9167–9170
26. Snatzke G (2000) Circular dichroism: an introduction. In: Berova N, Nakanishi K, Woody RW (eds) Circular dichroism – principles and applications, 2nd edn. Wiley-VCH, New York
27. Gottarelli G, Masiero S, Spada GP (1998) The use of CD spectroscopy for the study of the self-assembly of guanine derivatives. Enantiomer 3:429–438
28. Wen JD, Gray DN (2002) The Ff gene 5 single-stranded DNA-binding protein binds to the transiently folded form of an intramolecular G-quadruplex. Biochemistry 41:11438–11448
29. Esposito V, Randazzo A, Piccialli G, Petraccone L, Giancola C, Mayol L (2004) Effects of an 8-bromodeoxyguanosine incorporation on the parallel quadruplex structure $[d(TGGGT)]_4$. Org Biomol Chem 2:313–318
30. Uesugi S, Ikehara M (1977) Carbon-13 magnetic resonance spectra of 8-substituted purine nucleosides. Characteristic shifts for the syn conformation. J Am Chem Soc 99:3250–3253
31. Jordan F, Niv H (1977) C8-amino purine nucleosides. A well-defined steric determinant of glycosyl conformational preferences. Biochim Biophys Acta 476:265–271
32. Uesugi S, Ohkubo M, Urata H, Ikeara M, Kobayashi Y, Kyogoku Y (1984) Ribooligonucleotides, r(C-G-C-G) analogues containing 8-substituted guanosine residues, form left-handed duplexes with Z-form-like structure. J Am Chem Soc 106:3675–3676
33. Virgilio A, Esposito V, Randazzo A, Mayol L, Galeone A (2005) 8-Methyl-2'-deoxyguanosine incorporation into parallel DNA quadruplex structures. Nucleic Acids Res 33:6188–6195

34. Esposito V, Virgilio A, Pepe A, Oliviero G, Mayol L, Galeone A (2009) Effects of the introduction of inversion of polarity sites in the quadruplex forming oligonucleotide TGGGT. Bioorg Med Chem 17:1997–2001
35. Petraccone L, Duro I, Randazzo A, Virno A, Mayol L, Giancola C (2007) Biophysical properties of quadruplexes containing two or three 8-bromodeoxyguanosine residues. Nucleosides Nucleotides Nucleic Acids 26:669–674
36. Esposito V, Virgilio A, Randazzo A, Galeone A, Mayol L (2005) A new class of DNA quadruplexes formed by oligodeoxyribonucleotides containing a 3'-3' or 5'-5' inversion of polarity site. Chem Commun 31:3953–3955
37. Virno A, Zaccaria F, Virgilio A, Esposito V, Galeone A, Mayol L, Randazzo A (2007) Molecular modelling studies of four stranded quadruplexes containing a 3'-3' or 5'-5' inversion of polarity site. Nucleosides Nucleotides Nucleic Acids 26(8–9):1139–1143
38. Karsisiotis AI, Hessari NM, Novellino E, Spada GP, Randazzo A, Webba da Silva M (2011) Topological characterization of nucleic acid G-quadruplexes by UV absorption and circular dichroism. Angew Chem Int Ed Engl 50(45):10645–10648
39. Hu LY, Lim KW, Bouaziz S, Phan AT (2009) Giardia telomeric sequence $d(TAGGG)_4$ forms two intramolecular G-quadruplexes in K^+ solution: effect of loop length and sequence on the folding topology. J Am Chem Soc 131:16824–16831

Top Curr Chem (2013) 330: 87–110
DOI: 10.1007/128_2012_335

Published online: 3 July 2012

Molecular Crowding and Hydration Regulating of G-Quadruplex Formation

Daisuke Miyoshi, Takeshi Fujimoto, and Naoki Sugimoto

Abstract Intracellular space is highly crowded with soluble and insoluble biomolecules that range from large polymers, such as proteins and nucleic acids, to small molecules, including metabolites and metal ions. It is therefore of interest to understand the effects of molecular crowding on the structure, stability, and function of biomolecules. Moreover, molecular crowding is observed not only intracellularly but also in the extracellular matrix and under the conditions used in in vitro biotechnological and nanotechnological processes. However, most biochemical studies of biomolecules are performed under dilute conditions. Recent studies have demonstrated critical effects of molecular crowding on nucleic acids. In the present study we discuss how molecular crowding affects the properties of G-quadruplexes as well as other non-canonical nucleic acid structures.

Keywords Hydration · Molecular crowding · Structural polymorphism · Thermodynamics · Water activity

Contents

1 Introduction 88
2 Molecular Crowding 88
2.1 Molecular Crowding In Vivo 89
2.2 Molecular Crowding In Vitro 90

D. Miyoshi (✉) and N. Sugimoto (✉)
FIRST (Faculty of Frontiers of Innovative Research in Science and Technology), Konan University, 7-1-20 Minatojima-Minamimachi, Chuo-ku, Kobe 650-0047, Japan

FIBER (Frontier Institute for Biomolecular Engineering Research), Konan University, 7-1-20 Minatojima-Minamimachi, Chuo-ku, Kobe 650-0047, Japan
e-mail: miyoshi@center.konan-u.ac.jp; sugimoto@konan-u.ac.jp

T. Fujimoto
FIRST (Faculty of Frontiers of Innovative Research in Science and Technology), Konan University, 7-1-20 Minatojima-Minamimachi, Chuo-ku, Kobe 650-0047, Japan

2.3 Molecular Crowding Reagents 91
3 Molecular Crowding Effects on Non-canonical Structures of Nucleic Acids 92
3.1 DNA Triplexes 93
3.2 DNA Junctions 93
3.3 RNA Structures and Functions 93
4 Molecular Crowding Effects on G-Quadruplex Conformations 95
4.1 G-Quadruplex Conformations Under Molecular Crowding Conditions 96
4.2 Molecular Crowding Induced by Different Cosolutes 98
4.3 Duplex–Quadruplex Competitions 99
5 Molecular Crowding Effects on the Thermal Stability of G-Quadruplexes 100
5.1 Thermodynamics of G-Quadruplexes Under Molecular Crowding Conditions 100
5.2 Hydration of G-Quadruplexes 102
6 Conclusions and Perspectives: Making G-Quadruplexes More Canonical with Molecular Crowding 104
References 105

1 Introduction

One of the ultimate goals of biochemical and biophysical studies is to reveal how biomolecules participate in fundamental biological processes in situ, i.e., inside living cells. Intracellular space is highly crowded with soluble and insoluble biomolecules that range from large polymers, such as proteins and nucleic acids, to small molecules, including metabolites and metal ions. Surprisingly, the total molecular concentration within cells reaches 400 mg mL^{-1} and biomolecules occupy up to 40% of the total intracellular space (Fig. 1) [1–4], creating conditions of molecular crowding. It is therefore of interest to understand the effects of molecular crowding on the structure, stability, and function of biomolecules. Molecular crowding has been demonstrated to affect the properties of biomolecules through effects on the thermodynamics and kinetics of macromolecular association and dissociation [5–7]. However, most biochemical studies are performed under non-physiological dilute conditions [8, 9], in contrast to other cellular environmental factors, such as temperature, pH, ion species and concentrations, and redox potential, which are adjusted during biochemical experiments to match physiological conditions. Since there are excellent reviews regarding molecular crowding effects on the structure and function of proteins and nucleic acids [1–6, 8–10], we discuss herein how molecular crowding affects the properties of G-quadruplexes and other non-canonical nucleic acid structures.

2 Molecular Crowding

Molecular crowding conditions are observed both intracellularly and in the extracellular matrix and also in in vitro biotechnological and nanotechnological processes. In this section we demonstrate how molecular crowding is critical in both in vivo and in vitro environments.

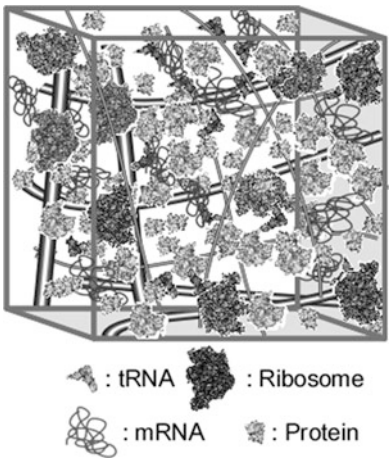


Fig. 1 Schematic illustration of molecular conditions within a living cell that is filled up with biomolecules such as tRNA, ribosomes, mRNA, and protein. Note that there are many small molecules in addition to these large biomolecules in a living cell

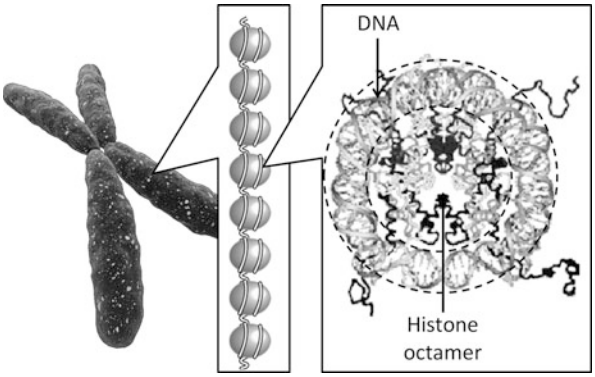


Fig. 2 Hierarchical structure of a chromosome. Nucleosomes (*right*), which are composed of a histone octamer and a DNA strand, form the beads-on-a-string structure that is shown in the *center*. A higher-order structure of nucleosomes makes a chromosome structure (*left*)

2.1 Molecular Crowding In Vivo

Molecular crowding occurs in both the nucleus and the cytoplasm where diagnostic target molecules exist and to where therapeutic molecules are targeted. Eukaryotic nuclei, in which DNA is present, are heterogeneous and contain a variety of subnuclear structures such as the nucleolus, splicing-factor compartments, Cajal bodies, promyelocytic leukemia bodies, replication factors, and transcription factors [11, 12]. In the nucleus, DNA is highly condensed and is packed with nuclear components to form chromosomes. DNA exists in a semi-ordered structure in these nuclear chromosomes, where it is wrapped around histones forming a composite material called chromatin (Fig. 2) [13, 14]. The nucleosome core

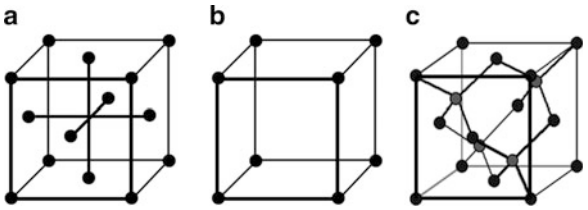


Fig. 3 Structure of (**a**) face-centered, (**b**) simple, and (**c**) diamond lattices. Atoms in a lattice are shown as *black circles*

consists of about 146 dsDNAs wrapped in left-handed superhelical turns around four identical pairs of proteins known as the histone octamer. DNA is then further packaged into metaphase chromosomes. Histone proteins regulate nucleosome assembly, DNA condensation, and DNA flexibility [15, 16]. Thus, the inside of the cell nucleus is crowded by molecules such as DNA and histones as well as by RNA, proteins, and small cosolutes.

Visualization of the cytoplasm, where RNA is present, by electron microscopy revealed the actin cytoskeleton, which was reconstructed without prior removal of membranes or extraction of soluble proteins [17, 18]. By using this technique, single macromolecular assemblies with distinct shapes, such as the proteasome and ribosomes, can be identified at a resolution of nanometers in an unperturbed cellular environment. Moreover, the cytoplasm of a living cell consists of a large number of additional molecules such as nucleic acids, proteins, carbohydrates, and lipids.

Therefore, the biomolecules occupy 20–40% of the total volume of the intracellular environment. The total concentration of intracellular biomolecules is estimated to be 300–400 mg mL^{-1}, and the concentration is even higher in the mitochondrial matrix [1, 19–21]. The crowded nature of the cell interior can be visualized by comparison with lattice structures. In face-centered cubic lattices and simple cubic lattices, the atomic packing factors, which are the fractions of the volumes of crystal structures that are occupied by atoms, are 0.74 and 0.52, respectively (Fig. 3). The atomic packing factor of a diamond crystal is just 0.34, which is comparable with the fractional volume commonly occupied by biomolecules in living cells. Molecular crowding in living organisms involves high concentrations not only of macromolecules but also of small molecules. The concentration of such small molecules, which act as osmolytes, can reach the molar range. The most extreme example in mammals is the kidney medulla cells, which contain a urea concentration of up to 5.4 mol L^{-1} [22], which corresponds to 30% of the cell by mass. Thus, not only macromolecules but also small organic molecules are indispensable for maintaining homeostasis in all living organisms [23, 24].

2.2 *Molecular Crowding In Vitro*

It is obvious that molecular reactions that are designed to take place on a material surface in pharmaceutical and medicinal devices also proceed under conditions of high molecular crowding. Typical in vitro molecular crowding can be observed at

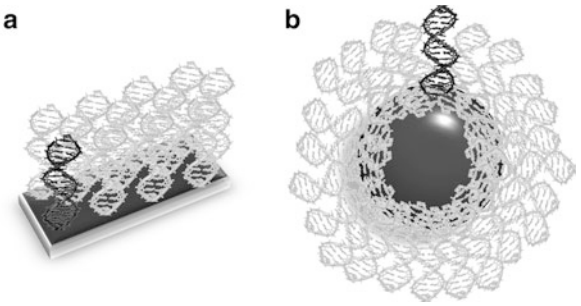


Fig. 4 DNA strands attached to (**a**) an immobilized surface and (**b**) an immobilized nanoparticle. For clarity, all DNA strands except one are shown in *gray*

liquid–solid interfaces, such as when a biomolecule is immobilized on a solid surface as occurs in analytical methods such as affinity chromatography, enzyme-linked immunosorbent assay (ELISA), and surface plasmon resonance (SPR). In such biosensing techniques, biomolecules are directly or indirectly attached to solid supports via covalent or non-covalent interactions (Fig. 4). When a DNA strand is directly immobilized, it is either surrounded by identical DNA strands or is on a self-assembled monolayer of an organic polymer that prevents non-specific adsorption of molecules onto the solid support. In the case of the direct immobilization of DNA strands onto gold nanoparticles, the density of the DNA strands was reported to be around 50 pmol cm^{-2} [25–27], resulting in around 3×10^{13} DNA strands cm^{-2}. Since the diameter of DNA is 2 nm, the DNA strands cover around 90% of the surface, indicating that there is almost no available free surface after hybridization of the DNAs with their target complementary DNA strands. This demonstrates the molecular crowding conditions of immobilized biomolecules. It is therefore possible that the hybridization process that occurs at the solution–solid interface is different from the process occurring in a dilute solution, and that this difference is at least partly due to molecular crowding.

In nanodevices, the distinctive chemical conditions at the solution-solid interface further affect the properties of the whole system. A nanopore, whose diameter generally ranges from several to several ten of nanometers, can be created by a membrane protein or by a combination of lithography and ultrathin materials such as silicon and graphene [28–30]. At the nanoscale level, the chemical condition at the liquid–solid interface that is not observed in bulk solution dominates the characteristics of the whole solution. Since such nanodevices have been applied to the detection, separation, and purification of biomolecules [31], it is essential to investigate the biomolecular properties at the solution–solid interface where the surface and immobilized biomolecules create molecular crowding conditions.

2.3 Molecular Crowding Reagents

Because it is difficult to investigate quantitatively nucleic acid interactions in a living cell and on material surfaces, in vitro experimental systems that use a

synthetic cosolute have been widely used to study biomolecular reactions under non-aqueous conditions. Such crowding cosolutes should meet the following criteria [10]: (1) they should be basically inert so that there is no chemical interaction between the target molecule and the crowding cosolute; (2) they should easily dissolve in water to stimulate molecular crowding; (3) in the case of large cosolutes, different polymer sizes should be available; and, (4) in the case of small cosolutes, different chemical properties should be available. Inert cosolutes in particular are convenient for the study of molecular environments. Thus neutral and highly water-soluble molecules are extremely useful cosolutes for quantitative studies. However, chemical interactions including electrostatic interactions between the target molecule and cosolutes should be taken into account when investigating intracellular biomolecular function [32, 33]. Poly(ethylene glycol) (PEG) and polysaccharides are often used to mimic molecular crowding conditions since they are inert to most biomolecules. These polymers can be dissolved in water to a relatively high concentration, and different polymer sizes are commercially available. Other small cosolutes that are useful for mimicking molecular crowding conditions are alcohols, glycols, amino acids, acetonitrile, trimethylamine *N*-oxide, and betaines [34–36]. Moreover, proteins such as chymotrypsin, albumin, hemoglobin, and lysozyme, as well as synthetic polymers such as poly(vinylpyrrolidone)s, are also utilized as crowding cosolutes to mimic better the chemical conditions inside living cells [37].

3 Molecular Crowding Effects on Non-canonical Structures of Nucleic Acids

The critical role of the nucleic acids, DNA and RNA, is to store and process the genetic information that includes all of the information that is necessary for life. In 1953, Watson and Crick demonstrated that DNA forms a double helix via Watson–Crick base pairs. Half a century after the discovery of the double helix, the Human Genome Project (HGP) identified the almost 3.2 billion base pairs in the entire human genome. The HGP also showed that repetitive DNA sequences, including dinucleotide and trinucleotide repeats, as well as telomeric and centromeric sequences, are widely distributed in the human genome. Most of this repetitive DNA can potentially fold into noncanonical structures. From this perspective, clarifying not only the primary structure (sequence) of nucleic acids but also thermodynamic and kinetic analyses of the higher-order structures and structure-function relationships of nucleic acids are essential for understanding the reactions that underlie biological processes and diseases. Thus, molecular crowding effects on the non-canonical structures of nucleic acids are of interest for a broad range of research fields. In this section, we will introduce the effects of molecular crowding on some of the non-canonical structures of nucleic acids.

3.1 DNA Triplexes

Spink and Chaires as well as Goobes and Minsky demonstrated that the Hoogsteen base pairs in a DNA triplex are generally stabilized by molecular crowding, whereas Watson–Crick base pairs are destabilized [38, 39]. These studies also suggested that the thermodynamics of DNA duplexes and triplexes are regulated by DNA hydration. Minsky et al. further estimated the thermodynamic parameters of the triplex formation of $T_{18}{\cdot}A_{20}{\cdot}T_{20}$ and found that stabilization of the DNA triplex by PEG was driven by a large negative enthalpy change that exceeded the unfavorable entropy change [40]. Based on the parameters measured under conditions of various salt concentrations and temperatures, they proposed that alterations in environmental conditions could be effectively compensated for by crowding cosolutes; this compensation provides a mechanism for adaptation to, and buffering against, unfavorable conditions such as unfavorable ionic strength or temperature [40].

3.2 DNA Junctions

Junctions in nucleic acid structures arise when three or more helices meet at a single point (Fig. 5a). These branched junctions are important intermediates in many biological functions and play important roles in many cellular processes [41]. Three-way junctions (TWJs) are the simplest type of junction and are used as a model system to gain insight into other complex and multi-branched junction structures [42]. TWJs create a unique electrostatic environment near the junction point due to the close proximity of opposing charges. The effects of molecular crowding on the formation of TWJs of DNA have been systematically studied [43]. TWJs, consisting of the junction point and three helical duplex arms, were destabilized by molecular crowding both in the absence and presence of 5 mM Mg^{2+}. However, the difference in the number of water molecules that were taken up by the TWJ as a whole and by each helical arm demonstrated that water molecules were released from the junction point. This dehydration upon the junction point formation suggested the stabilization of the junction point under the molecular crowding conditions. In fact, molecular crowding induced the structural transition from a bimolecular duplex to the intramolecular TWJ, even in the absence of Mg^{2+} (Fig. 5b). This finding demonstrates that intracellular conditions, under which water activity decreases and hydration becomes less favorable, facilitate the formation of TWJ structures rather than bimolecular duplex structures.

3.3 RNA Structures and Functions

There have been several studies of molecular crowding effects on high-order RNA structures and their catalytic activity. It was reported that methanol stabilized

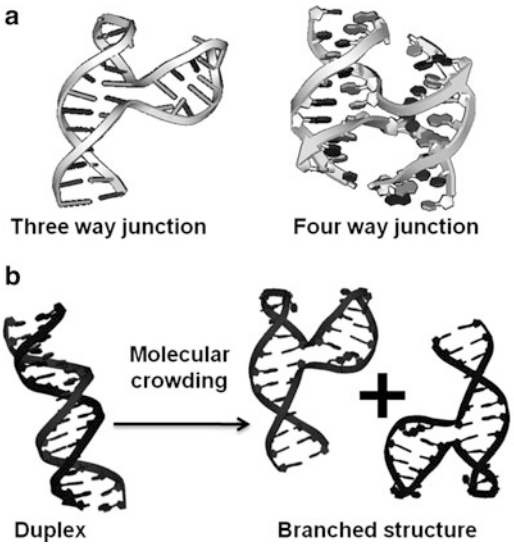


Fig. 5 (**a**) Three-dimensional structure of three- and four-way junctions. (**b**) Schematic illustration of the molecular crowding effect on the equilibrium between an intermolecular duplex and an intramolecular three-way junction

a 58-nt RNA structure but not RNA secondary structure [44]. Addition of the osmolyte trimethylamine-oxide (TMAO) enhanced the efficiency of ribosome reconstitution by up to 100-fold, providing a substantially improved system for the in vitro analysis of mutant ribosomes [45]. These findings demonstrated that cosolutes play an important role in stabilizing high-order RNA structures and their interactions with other biomolecules. This effect may be due to the stabilizing effect of TMAO on RNA tertiary structure. Indeed, TMAO can counteract the denaturing effects of urea on tRNA tertiary structure [46]. In order to understand the basis for the effects of TMAO on RNA structures, Draper and coworkers quantified the TMAO-induced stabilization of RNAs. They suggested that the formation of RNA tertiary structure is accompanied by substantial dehydration of the phosphate groups, and that TMAO affects this process by reducing the energetic penalty associated with this dehydration [47]. These results were supported by molecular dynamic simulations of a 22-nt RNA hairpin in aqueous TMAO solution [48]. Further reports of molecular simulations indicated that molecular crowding enhances the thermodynamic stability of a pseudoknot in human telomerase RNA, leading to a structural switch from a hairpin to a pseudoknot of human telomerase RNA [49].

It was reported that the RNA cleavage activity of a hairpin ribozyme increases in the presence of PEG 400 or dextran 10000 [50, 51]. The RNA compaction accompanied by water release that occurs during the folding of the hairpin ribozyme into the active conformation was proposed to be responsible for the enhanced cleavage activity under molecular crowding conditions. It was also found that the formation of a tertiary RNA structure of the GAAA tetraloop-receptor RNA tertiary motif that was observed in the group I ribozyme domain was favored by molecular crowding induced by higher molecular weight PEG and dextran but was not

induced by lower molecular weight sucrose and glycerol [52]. Small angle X-ray scattering experiments showed that PEG favors more compact RNA structures and folding at low Mg^{2+} concentrations because of the excluded volume effects of molecular crowding [53]. Although the presence of the cosolutes decreased the thermal stability of the ribozyme stem (duplex) regions, the structure as a whole was stabilized by molecular crowding. More importantly, it was found that molecular crowding induced by PEG decreased the metal ion concentration that was required for the catalytic activity of a hammerhead ribozyme, facilitated catalytic turnover, and activated a ribozyme that was in inactive form [54]. This ribozyme also showed efficient activity in a solution at physiological salt concentration only under osmotic pressure induced by added cosolutes. The fact that the ribozyme functioned efficiently under physiological conditions implies that ribozymes have the potential to work efficiently within a living cell. These results indicate that molecular crowding is critical not only to maintain RNA structures but also to maintain their functions in living cells where there exist many biomolecules that can destabilize RNA structures.

4 Molecular Crowding Effects on G-Quadruplex Conformations

G-quadruplexes have a unique polymorphic nature [55–59]. The polymorphism of G-quadruplex conformation is induced and regulated not only by nucleotide sequence but also by chemical conditions. Since a G-quadruplex is stabilized by cation coordination to the center of the G-quartet, the effects of cations on the conformation and thermal stability of G-quadruplexes have been studied [60–69]. Using NMR, Wang and Patel were the first to show that the human telomeric DNA sequence, $AG_3(TTAG_3)_3$, folded to form an antiparallel G-quadruplex in the presence of Na^+ [70]. Subsequently, Parkinson et al., using X-ray crystallography, reported the formation of a parallel G-quadruplex of $AG_3(TTAG_3)_3$ in the presence of K^+ [71]. There are now at least five types of intermolecular G-quadruplexes known for human telomeric DNA sequences that contain TTAGGG repeats. NMR and chemical modification studies have shown the formation of mixed G-quadruplexes [(3+1) G-quadruplexes or hybrid G-quadruplexes] that involve Form 1 and Form 2 [72–74]. These studies have demonstrated that chemical conditions as well as nucleotide sequences should be taken into account when determining native G-quadruplex conformations in living cells. In addition to the effects of cation species and their concentration, it is now widely accepted that molecular crowding critically affects the structural polymorphism of G-quadruplexes. In this section we will discuss how molecular crowding regulates the conformation and the thermal stability of G-quadruplexes.

Table 1 G-quadruplex conformations under molecular crowding conditions

Ref.	Sequence	Cosolute	Cation	Conformation
[75]	$G_4T_4G_4$ & $(G_4T_4)_3$ G_4	2 M PEG	100 mM Na^+	Parallel
		2 M Putrescine		Antiparallel
[76]	$A(G_3T_2A)_3$	1.4 M PEG	185 mM Na^+	Antiparallel
			185 mM K^+	Mixed[a]
[77]	$(G_3T_2A)_3$ G_3	40% PEG	100 mM Na^+	Antiparallel
	$(G_4T_2)G_4$	40% PEG		Parallel (G-wire)
[78]	$(G_3T_2A)_3$ G_3	40% PEG	150 mM Na^+	Antiparallel
			150 mM K^+	Parallel
[79]	$(G_3T_2A)_3$ G_3	2 M PEG	None	Parallel
[80]	$(G_3T_2A)_3$ G_3[b] and $A(G_3T_2A)_3$ G_3	Ethanol	150 mM K^+	Mixed and parallel[c]
[81]	$AG_3(T_2AG_3)_3$	50% Acetonitrile	25 mM K^+	Parallel[d]
		300 mg mL^{-1} BSA		Unassigned
[82][e]	Human telomere	Neutral cosolutes		Parallel
[83]	Human telomere[f]	40% PEG	K^+	Parallel
		Egg extract[g]		Not parallel
[84]	TBA	40% PEG	100 mM Na^+ or K^+	Antiparallel

[a]CD spectra was typical for the mixed conformation
[b]Longer human telomere DNA sequences were also studied
[c]The conformation depended on DNA concentration
[d]CD spectra was typical for the parallel but NMR spectra was not (see main text)
[e]Human telomere sequences with various flanking sequences were studied with PEGs, dimethyl sulfoxide, ethanol, and acetonitrile. Higher concentrations of PEG 200 induced a high-order structure
[f]Human telomere DNA sequences with various flanking sequences were studied
[g]Egg extract indicates *Xenopus laevis* egg extract

4.1 G-Quadruplex Conformations Under Molecular Crowding Conditions

The effect of molecular crowding on G-quadruplex conformations was first studied using the *Oxytricha nova* telomeric DNA sequence $dG_4T_4G_4$, in the presence of Na^+ (Table 1). Molecular crowding conditions were created by use of both the neutral cosolutes PEG (MW = 300) and glycerol and the positively charged cosolutes putrescine, cadaverine, and spermidine [75]. It was found that molecular crowding induced by PEG induced a conformational transition from an antiparallel to a parallel conformation of the $d(G_4T_4G_4)$ and $d(G_4T_4)_3$ G_4 G-quadruplexes (Fig. 6a), whereas molecular crowding using the cationic cosolutes did not alter the antiparallel G-quadruplex conformation. These results suggested that the conformational switch of G-quadruplexes is regulated by molecular crowding.

In 2005, Chaires and coworkers reported molecular crowding effects on the G-quadruplex conformation of the human telomeric DNA sequence $dA(G_3TTA)_4$ in the presence of Na^+ and K^+ [76]. Circular dichroism and fluorescence quenching

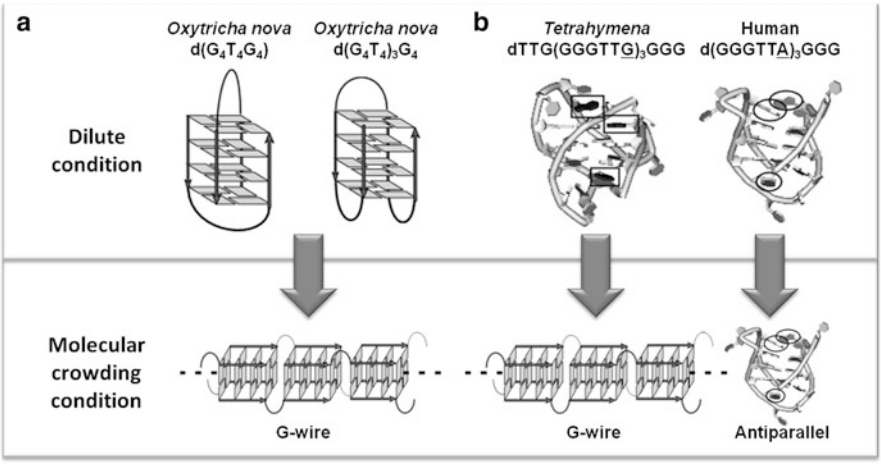


Fig. 6 Molecular crowding effects on the conformations of (**a**) *Oxytricha nova* telomere DNA and (**b**) *Tetrahymena* and human telomeric DNAs. Molecular crowding induces the parallel-stranded G-quadruplex (G-wire) conformation of all of the telomeric DNA sequences except for human telomeric DNA

studies showed dramatic changes in the G-quadruplex conformation under molecular crowding conditions in the presence of K^+ but not in the presence of Na^+. These results, as well as the results obtained for the *O. nova* telomeric DNA, indicated that molecular crowding effects on G-quadruplex conformations are dependent on the nature of the coexisting cation and on the nucleotide sequence. To study the polymorphic nature of DNA G-quadruplexes under molecular crowding conditions, the effects of the telomeric DNA sequence on the intermolecular and intramolecular G-quadruplexes formed from *Tetrahymena*, human, and *O. nova* telomeric sequences were studied under dilute and molecular crowding conditions in the absence and presence of various cosolutes [77]. The intermolecular and intramolecular G-quadruplex structures formed from *Tetrahymena* telomeric DNA formed very long, well-ordered G-wires in the presence of cosolutes in Na^+ solution. Although this structure is adopted by many parallel-stranded telomeric sequences, the intermolecular and intramolecular G-quadruplex structures formed from human telomeric sequences remained as compact antiparallel G-quadruplex structures under these conditions (Fig. 6b). These results demonstrate that a single G to A base replacement in the loops of a G-quadruplex leads to a dramatically different structure under molecular crowding conditions in the presence of Na^+. Conversely, the parallel conformation of a G-quadruplex formed from human telomeric DNA was observed in a K^+-containing solution under molecular crowding conditions [78]. More complex behavior of the human telomeric DNA under molecular crowding conditions has also been reported: in the presence of K^+, molecular crowding using ethanol induced a conformational transition of human telomeric DNA from the antiparallel to the parallel conformation via a mixed conformation [80]. Such structural conversion of the human telomeric DNA by molecular

crowding was also observed in the absence of a monovalent cation [79] and, furthermore, was observed for long telomeric sequences containing two G-quadruplex units [81]. Heddi and Phan used NMR to study how molecular crowding affects the human telomeric DNAs that had folded into the four different conformations under dilute conditions in the presence of K^+ [82]. The NMR spectra clearly demonstrated that the four different G-quadruplex conformations were all converted to the parallel G-quadruplex conformation under molecular crowding conditions induced by PEG 200 in the presence of K^+. Notably, this parallel conformation was almost identical to the parallel conformation observed by X-ray crystallography under dilute conditions with K^+. Furthermore, the addition of other the neutral cosolutes PEG 8000, dimethyl sulfoxide, ethanol, and acetonitrile also induced a similar conformation as that induced by PEG 200. These NMR results indicate that molecular crowding simplifies the polymorphic nature of G-quadruplexes, depending on the nucleotide sequence of the G-quadruplex.

4.2 Molecular Crowding Induced by Different Cosolutes

Although molecular crowding generally favors the parallel conformation of a G-quadruplex, the above results indicated that the molecular crowding effects on G-quadruplex conformation are highly dependent, not only on the nucleotide sequence of the G-quadruplex and the nature of the co-existing cation species but also on the cosolute species. From this perspective, a vertebrate telomeric DNA G-quadruplex was studied under molecular crowding conditions within a *Xenopus laevis* egg extract as well as in the presence of PEG 200 and Ficoll70 [83]. The conformation of the vertebrate telomeric DNA in the *X. laevis* egg extract or in Ficoll was different from that observed in the presence of PEG. Based on these results, the authors stated that PEG should not be used to mimic molecular crowding conditions. However, the usage of PEG to induce molecular crowding conditions can provide important information for evaluation of quantitative parameters including thermodynamic parameters of G-quadruplex structure and the hydration state of G-quadruplexes, as will be discussed below. Moreover, Trent and coworkers, using acetonitrile as a non-hydrogen-bonding dehydrating agent, performed CD and NMR studies of the conformation of human telomeric DNA in K^+ and Na^+ solutions [85]. They showed that, although the CD spectra obtained using acetonitrile were similar to those obtained using PEG 400, the NMR spectra obtained using these reagents were not similar to each other, indicating that assignment of G-quadruplex conformations based only on CD spectra is not sufficient. Interestingly, the NMR spectrum obtained under the molecular crowding condition induced by bovine serum albumin was similar to that obtained under the dilute condition.

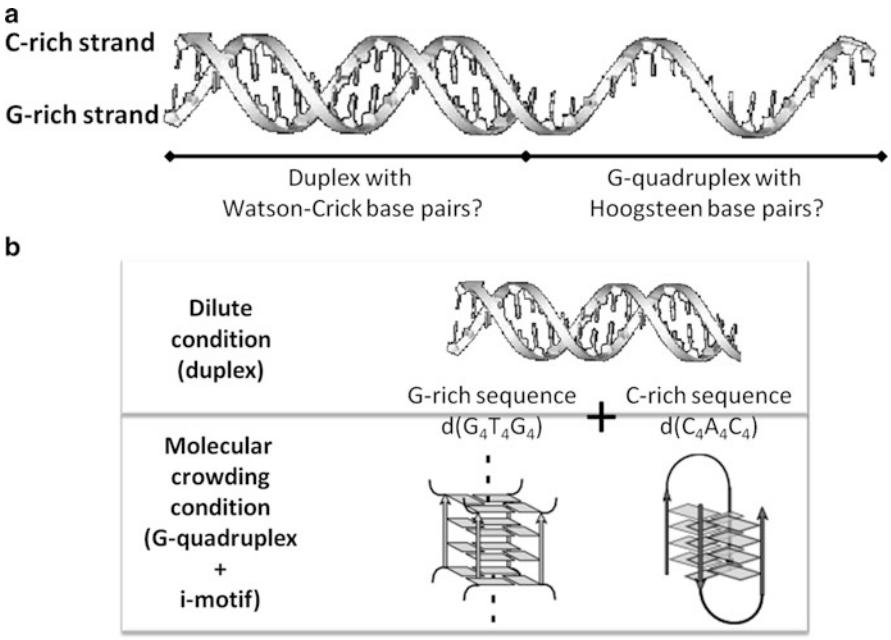


Fig. 7 (**a**) Schematic illustration of a telomeric region composed of G-rich and C-rich strands. (**b**) Structures of the telomeric DNAs under dilute and molecular crowding conditions. Under dilute and molecular crowding conditions the 1:1 mixture of the G-rich and C-rich strands folds into a duplex and quadruplexes, respectively

4.3 Duplex–Quadruplex Competitions

It has been considered that most telomeric DNAs fold into a duplex with Watson–Crick base pairing (Fig. 7a). However, it remains unclear whether duplex formation is dominant in the G-rich/C-rich regions since these G-rich and C-rich regions can individually form quadruplex structures [86]. Thus, molecular crowding effects on duplex–quadruplex competition are of interest in terms of predicting the native structure of such regions. It was demonstrated that a duplex formed with G- and C-rich DNAs under a dilute condition in the presence of Na^+ was dissociated upon molecular crowding and that the G- and C-rich DNA sequences folded into individual quadruplexes (Fig. 7b) [87]. Tan and coworkers also reported that molecular crowding induced telomere G-quadruplex formation in a salt-deficient solution and in a K^+-containing solution and that it enhanced the competition between G-quadruplex formation and duplex formation [88, 89]. Kinetic studies of G-quadruplex formation from a duplex with the complementary sequence showed that the formation rate of a G-quadruplex depends on the PEG and the K^+ concentration [90]. A quantitative analysis showed that, when 30 nmol L^{-1} of G-rich and C-rich human telomeric DNAs were mixed together in a 100 mM K^+ solution, the amount of the G-quadruplex formed was 17.6 and 23.4 nmol L^{-1} in the

absence and presence of 10% ethylene glycol, respectively [91]. As shown using these short oligonucleotides, molecular crowding made it possible for a G-rich sequence to form a stable G-quadruplex in a long double-stranded DNA [92]. These results indicate that molecular crowding is essential for G-quadruplex formation in the G-rich/C-rich double stranded region. In such regions, the polymorphic nature of the G-quadruplex can be induced by molecular crowding in vivo. These results also illustrate the difficulty of correctly extrapolating the structures of not only telomeric DNA sequences but also of many other biomolecules in vivo based on their structures in vitro. Thus, molecular crowding as well as other cellular environmental factors [93–96] play critical roles in the structure of a biomolecule and analysis of their effects is useful for understanding biomolecular behaviors in vivo and for efficient drug and ligand targeting by biomolecules.

5 Molecular Crowding Effects on the Thermal Stability of G-Quadruplexes

As discussed above, molecular crowding stabilizes non-canonical structures of nucleic acids that involve DNA triplexes, DNA junctions, and RNA tertiary structures, and strongly stabilizes non-canonical G-quadruplexes. Moreover, molecular crowding creates and regulates the conformational changes between different types of G-quadruplexes, and between a G-quadruplex and a duplex. In this section we will discuss how molecular crowding stabilizes G-quadruplex structures, resulting in dynamic conformational transitions.

5.1 Thermodynamics of G-Quadruplexes Under Molecular Crowding Conditions

Table 2 summarizes the thermodynamic parameters of G-quadruplex formations under dilute and molecular crowding conditions. It is widely accepted that molecular crowding significantly stabilizes G-quadruplexes. In order to show the differences in molecular crowding effects on a non-canonical G-quadruplex and a canonical duplex, the thermodynamic parameters of the intramolecular antiparallel G-quadruplex formed by a thrombin binding aptamer (TBA) and those of an intramolecular antiparallel hairpin-looped duplex were evaluated under molecular crowding conditions induced by various neutral cosolutes [84]. It was quantitatively demonstrated that the antiparallel G-quadruplex is stabilized under molecular crowding conditions at various concentrations of PEG 200 but that these conditions destabilize the duplex. The melting temperature (T_m) for 5 μmol L^{-1} of the antiparallel G-quadruplex increased from 54.1 °C to 58.7 °C as the PEG 200 concentration was increased from 0 to 40 wt% in a K^+-containing solution. In contrast to the

Table 2 G-quadruplex stabilities under molecular crowding conditions[a]

Ref.	Sequence	Cosolute	ΔT_m (°C)	$\Delta\Delta G^{\circ}$[b] (kcal mol^{-1})	$\Delta\Delta H^{\circ}$ (kcal mol^{-1})	$T\Delta\Delta S^{\circ}$[b] (kcal mol^{-1})
[84]	TBA[c]	40% PEG	4.6	−2.0	−11	−9
[97]	$G_3T_3G_3TG_3T_3G_3$	30% EG[d]	12	−3.5	−16	−12
	$G_3T_5G_3TG_3T_5G_3$		15	−2.4	−6.4	−3.8
[79][e]	$(G_3T2_A)_3\ G_3$	0.75 M PEG 200	11	−4.6	+39	+43
		0.75 M PEG 400	14	−2.1	+69	+71
[85]	$AG_3(T_2AG_3)_3$	50% Acetonitrile	19	−5.9	−6.1	N.D.
[98]	*bcl*-2 DNA	40% PEG	26	−4.5	−12	−7.6
	bcl-2 RNA		20	−3.6	−6.6	−3.0

[a]The molecular crowding effects are evaluated as the follow: (Parameter under the crowding condition) − (Parameter under the dilute condition)
[b]Thermodynamic parameters at 25 °C
[c]TBA indicates thrombin binding aptamer
[d]EG indicates ethylene glycol
[e]The parameters are the difference between 0.75 and 0.1 M cosolute because the experiments were carried out without cation

G-quadruplex, the T_m for 5 μmol L^{-1} of the duplex in the presence of K^+ decreased from 66.4 °C to 54.3 °C as the PEG 200 concentration was increased from 0 to 40 wt%. When the PEG 200 concentration was increased from 0 to 40 wt%, the values of ΔH° (enthalpy change), $T\Delta S^{\circ}$ (entropy change), and ΔG°_{25} (free energy change at 25 °C) of the antiparallel G-quadruplex decreased from −42.0 to −53.0 kcal mol^{-1}, from −38.5 to −47.5 kcal mol^{-1}, and from −3.5 to −5.5 kcal mol^{-1}, respectively. The difference in ΔG°_{25} induced by molecular crowding was −2.0 kcal mol^{-1}. On the other hand, the values of ΔH°, $T\Delta S^{\circ}$, and ΔG°_{25} for formation of the duplex increased from −81.5 to −75.8 kcal mol^{-1}, from −71.7 to −68.9 kcal mol^{-1}, and from −9.8 to −6.9 kcal mol^{-1}, respectively, with the same change in PEG 200 concentration. These changes indicated that promotion of the G-quadruplex formation by molecular crowding was enhanced by a favorable enthalpic contribution that exceeded an unfavorable entropic contribution. Conversely, the duplex destabilization was due to an unfavorable enthalpic contribution that exceeded a favorable entropic contribution. Such an enhancement in the thermodynamic stability of G-quadruplexes by molecular crowding has been reported for various G-rich nucleotide sequences. In particular, the G-quadruplex formed by the sequence $dG_3T_3G_3TG_3T_3G_3$ was stabilized 3.5 kcal mol^{-1} in ΔG°_{25} by 30% ethylene glycol in a K^+ solution, which corresponded to more than two orders of difference in the equilibrium constant [97]. Stabilization effects were also reported for the structures formed by the human telomeric DNA [79, 85] and by *bcl*-2 DNA [98]. It is noteworthy that the stabilization and destabilization of DNA structures do not depend on their structure but instead depend on the type of base pairs that are involved in their structure [99]. In addition, it was recently reported that the Hoogsteen base pairs in triplex and G-quadruplex DNA structures were stabilized not only by molecular crowding but also by a model peptide that mimicked histone H3, which is critical for the formation of a high-order complex

called a nucleosome that involves a DNA strand and histone octamers (Fig. 2) [100]. Since nucleosomes become further organized to form chromatin inside the eukaryotic cell nucleus, and since chromatin structure is dynamic and controls gene expression, the stabilization of G-quadruplexes and triplexes by a histone-mimicking peptide implies roles for such non-canonical structures in transcription. The stabilization of G-quadruplexes by molecular crowding supports both the formation of G-quadruplexes and the biological roles of G-quadruplexes within living cells.

5.2 *Hydration of G-Quadruplexes*

Water molecules play critical roles in generating and maintaining the structure, stability, and function of biomolecules. It is well known that many proteins unfold in non-aqueous solution and lose their functions in the absence of water. Nucleotides have a large number of hydration sites. The 11–12 hydration sites of a nucleotide are occupied by water molecules that directly interact with the nucleotide and form a primary hydration shell in which 8–9 water molecules per nucleotide are bound to the primary hydration shell [101]. These water molecules play fundamental roles in maintaining the structure, stability, and function of nucleic acids.

The use of the osmotic stress technique that uses an osmolyte as a cosolute provides information regarding biomolecular interactions with water molecules. Such analysis has revealed differences in the numbers of water molecules bound, $-\Delta n_w$, which are evaluated as follows: $-\Delta n_w = n_{w,folded} - n_{w,unfolded}$, where $n_{w,folded}$ and $n_{w,unfolded}$ indicate the numbers of water molecules bound to the folded and unfolded states, respectively. In general, an osmolyte that induces molecular crowding lowers the activity of water, a_w, and thus decreases the chemical potential of water. Therefore, cosolutes affect the equilibrium constant of biomolecular reactions that involve the association or dissociation of water molecules. Indeed, water molecules participate in most biochemical reactions. The dependency of water activity on the equilibrium constant K of a biomolecular reaction reflects the number of water molecules released upon the reaction Δn_w as presented by the equation $(\partial \log K/\partial \log a_W) = -\Delta n_w$, if other interactions between the cosolute and the nucleic acid, and between the cosolute and the water molecule, are negligible [102]. This technique can be applied to any equilibrium and parameters to evaluate how many water molecules participate in the reactions.

The osmotic stress method has been applied to the evaluation of the hydration state of G-quadruplexes. UV melting studies of the TBA in a K^+ solution at 0 and 40 wt% PEG 200 showed that molecular crowding with PEG 200 stabilizes the TBA G-quadruplex structure [84]. Thermodynamic parameters including the equilibrium constant K_{obs} can be evaluated based on the melting curves. Plots of ln K_{obs} vs. ln a_w can be drawn based on osmotic pressure measurements. Both of these K_{obs} have a linear relationship with ln a_w, and the slope of this line corresponds to the

Table 3 Hydration of G-quadruplexes and other DNA structures

Ref.	Sequence	Structure	Cation	Δn_w[a]
[84]	$G_2T_2G_2TGTG_2T_2G_2$	Antiparallel (chair)	100 mM K^+	4.5
			100 mM Na^+	4.0
	28 mer[b]	Antiparallel duplex	100 mM K^+	−3.4
			100 mM Na^+	−3.5
[103]	$G_2T_2G_2TGTG_2T_2G_2$	Antiparallel (chair)	100 mM K^+	0.53
	$G_3T_2G_3TGTG_3T_2G_3$	Antiparallel (chair)		1.1
	$TG_4AG_3TG_4AG_3TG_4A_2G_2$	Parallel		1.5
	$AG_3T_2AG_3T_2AG_3T_2AG_3$	Antiparallel		0.14
[104]	$G_2U_2G_2UGUG_2U_2G_2$	Antiparallel	100 mM K^+	0.6
	$G_2T_2G_2TTTG_2T_2G_2$			1.4
[97]	$G_3T_3G_3T_3G_3TG_3$	Parallel	100 mM K^+	0.27
	$G_3T_3G_3T_5G_3TG_3$	Mixed	100 mM K^+	0.14
[105]	$A(G_3T_2A)_3G_3T_2$	Mixed[c]	100 mM K^+	0.17
[98]	*bcl*-2 DNA	Parallel[d]	5 mM K^+	8.5

[a]The values of Δn_w are per nucleotide. The negative and positive Δn_w indicate hydration and dehydration, respectively, upon the structure formation

[b]The sequence is TCTTTCTCTTCTTTTTCGAAGAGAAAGA (underlined is loop region)

[c]Structure under the dilute condition is the mixed conformation

[d]Structure under the molecular crowding condition is the parallel conformation

number of water molecules released upon structure formation, showing that water molecules are released upon the formation of G-quadruplexes. In contrast, water molecules are taken up upon duplex formation. It was quantitatively calculated that 4.5 and 4.0 water molecules per nucleotide were released upon G-quadruplex formation. Conversely, in solutions of K^+ and Na^+, 3.4 and 3.5 water molecules per nucleotide were taken up upon duplex formation.

The numbers of water molecules released upon G-quadruplex formation have been reported for various sequences (Table 3). Marky and coworkers reported values of Δn_w for DNA aptamers, NHE-III, and a human telomere [103] in a K^+ solution. They calculated that 0.1–1.5 water molecules were released per nucleotide through G-quadruplex folding depending on the nucleotide sequence. It was proposed that these values result from (1) the release of structural water from the random coil state upon formation of the G-quadruplex, (2) the uptake of electrostricted water molecules by G-quadruplexes with a higher charge density, and (3) the release of electrostricted water from K^+ upon binding to the G-quadruplex core (G-quartet). Notably, they further separately evaluated Δn_w for the loop and G-quartet (G-quadruplex core) regions using various substitutions [104] and reported that $\Delta n_w = 13$ for the G-quartet and $\Delta n_w = -6 \sim +8$ for the remaining loop region. In addition, by comparison of the Δn_w values of not only G-quadruplexes but also of duplexes and triplexes, it was proposed that the opposing effects of molecular crowding on canonical structures (duplexes) and on non-canonical structures (G-quadruplexes and triplexes) were due to different behaviors of water molecules binding to the DNA strands [99]. Trent and coworkers further attempted to assess experimentally the role of steric crowding (excluded volume)

and hydration on the structure of a human telomeric G-quadruplex. Using BSA and acetonitrile to induce molecular crowding conditions, it was shown that the excluded volume effect on G-quadruplex conformation under molecular crowding conditions induced by acetonitrile is small and that hydration is the dominant factor for determination of the conformation and stability of a G-quadruplex [85]. Although the number of water molecules released upon G-quadruplex formation varied from 0.14 to 8.5 depending on the experimental conditions and procedure used as well as on the nucleotide sequence and the conformation of the G-quadruplex (Table 3), the dehydration of the G-quadruplex through structure formation is clearly observed, and this dehydration should stabilize the G-quadruplex under cell-mimicking conditions.

6 Conclusions and Perspectives: Making G-Quadruplexes More Canonical with Molecular Crowding

In the present review we have introduced and discussed how molecular crowding alters the conformation and thermodynamics of nucleic acids with structures ranging from a canonical duplex to a non-canonical G-quadruplex. Changes in nucleic acid structures and their stabilities affect various enzymes including nucleases, polymerases, telomerases, and helicases [78, 106–110]. Although the molecular crowding effects on enzyme functions are not yet sufficiently understood, it is now generally accepted that molecular crowding is one of the most important players in the regulation of various biological processes via stabilization or destabilization of nucleic acid structures and enhancement or suppression of nucleic acid functions. In particular, molecular crowding effects on G-quadruplexes should be taken into account when considering the biological functions of G-quadruplexes because of the polymorphic nature of their conformation, thermal stability, and dynamics. In fact, the properties of G-quadruplexes under molecular crowding conditions critically affect ligand and drug designs. Tan and coworkers studied molecular crowding effects on the functions of G-quadruplex ligands in terms of their effects on the affinity of ligand binding to G-quadruplexes and on the efficiency of these ligands for inhibition of telomerase activity [111]. They found that, under molecular crowding conditions, the ligands TMPyP4, BMVC, and Hoechst 33258 became significantly less effective, or lost the ability to stabilize the G-quadruplex and to inhibit telomerase activity. On the other hand, recent studies showed that TMpyP4 bound to G-quadruplex structures with higher affinity than to duplexes [112]. The binding of TMPyP4 to G-quadruplexes of human and *O. nova* telomeric DNAs under molecular crowding conditions has also been reported [113, 114]. Moreover, some G-quadruplex ligands that can bind to G-quadruplexes with high affinity and specificity under molecular crowding conditions have been found [115, 116]. It will therefore be necessary to determine how molecular crowding affects G-quadruplex ligand functions and how we can

rationally design G-quadruplex ligands so that they function under the conditions within living cells.

Molecular crowding effects on biomolecules are also important for various in vitro technologies, including drug delivery systems [117], the dispersion of carbon nanomaterials [118, 119], nanoparticle assembly [120, 121], response enhancement of electrochemical DNA sensors [122], DNA strand exchange [123], and the manipulation of single DNA molecules [124]. These studies clearly demonstrate that molecular crowding is a useful chemical stimulus for controlling the properties of nucleic acids toward in vitro applications. Since molecular crowding destabilizes canonical duplexes but stabilizes non-canonical structures especially G-quadruplexes, it will be possible to make G-quadruplex more canonical structure of nucleic acids with molecular crowding both in vivo and in vitro.

Acknowledgments This work was supported in part by Grants-in-Aid for Scientific Research, Scientific Research on Innovative Areas "Nanomedicine Molecular Science" (No. 2306), the "Strategic Research Foundation at Private Universities" (2009–2014), and the "Academic Frontier" project (2004–2009) from the Ministry of Education, Culture, Sports, Science and Technology, Japan, and the Hirao Taro Foundation of the Konan University Association for Academic Research. TF is a research fellow of Japan Society for the promotion of science.

References

1. Zimmerman SB, Trach SO (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. J Mol Biol 222:599–620
2. Parsegian VA, Rand RP, Rau DC (2000) Osmotic stress, crowding, preferential hydration, and binding: a comparison of perspectives. Proc Natl Acad Sci USA 97:3987–3992
3. Minton AP (2001) The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. J Biol Chem 276:10577–10580
4. Ellis RJ, Minton AP (2003) Join the crowd. Nature 425:27–28
5. Zimmerman SB, Minton AP (1993) Macromolecular crowding: biochemical, biophysical, and physiological consequences. Annu Rev Biophys Biomol Struct 22:27–65
6. Zhou H-X, Rivas G, Minton AP (2008) Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. Annu Rev Biophys 37:375–397
7. Lukacs GL, Haggie P, Seksek O, Lechardeur D, Freedman N, Verkman AS (2000) Size-dependent DNA mobility in cytoplasm and nucleus. J Biol Chem 275:1625–1629
8. Ellis RJ (2001) Macromolecular crowding: obvious but underappreciated. Trends Biochem Sci 26:597–604
9. Ellis RJ (2001) Macromolecular crowding: an important but neglected aspect of the intracellular environment. Curr Opin Struct Biol 11:114–119
10. Miyoshi D, Sugimoto N (2008) Molecular crowding effects on structure and stability of DNA. Biochimie 90:1040–1051
11. Lewis JD, Tollervey C (2000) Like attracts like: getting RNA processing together in the nucleus. Science 288:1385–1389
12. O'Brien TP, Bult CJ, Cremer C, Grunze M, Knowles BB, Langowski J, McNally J, Pederson T, Politz JC, Pombo A, Schmahl G, Spatz JP, van Driel R (2003) Genome function and nuclear architecture: from gene expression to nanoscience. Genome Res 13:1029–1041

13. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389:251–260
14. Richmond TJ, Davey CA (2003) The structure of DNA in the nucleosome core. Nature 423:145–150
15. Narlikar GJ, Fan H-Y, Kingston RE (2002) Cooperation between complexes that regulate chromatin structure and transcription. Cell 108:475–487
16. Bloomfield VA (1996) DNA condensation. Curr Opin Struct Biol 6:334–341
17. Beck M, Lucić V, Förster F, Baumeister W, Medalia O (2007) Snapshots of nuclear pore complexes in action captured by cryo-electron tomography. Nature 449:611–615
18. Medalia O, Weber I, Frangakis AS, Nicastro D, Gerisch G, Baumeister W (2002) Macromolecular architecture in eukaryotic cells visualized by cryoelectron tomography. Science 298:1209–1213
19. Srere PA (1981) Protein crystals as a model for mitochondrial matrix proteins. Trends Biochem Sci 6:4–7
20. Fulton AB (1982) How crowded is the cytoplasm? Cell 30:345–347
21. Goodsell DS (1991) Inside a living cell. Trends Biochem Sci 16:203–206
22. MacMillen RE, Lee AK (1967) Australian desert mice: independence of exogenous water. Science 158:383–385
23. Yancey PH, Clark ME, Hand SC, Bowlus RD, Somero GN (1982) Living with water stress: evolution of osmolyte systems. Science 217:1214–1222
24. Lang F, Busch GL, Ritter M, Völkl H, Waldegger S, Gulbins E, Häussinger D (1998) Functional significance of cell volume regulatory mechanisms. Physiol Rev 78:247–306
25. Akamatsu K, Kimura M, Shibata Y, Nakano S, Miyoshi D, Nawafune H, Sugimoto N (2006) A DNA duplex with extremely enhanced thermal stability based on controlled immobilization on gold nanoparticles. Nano Lett 6:491–495
26. Sato K, Hosokawa K, Maeda M (2003) Rapid aggregation of gold nanoparticles induced by non-cross-linking DNA hybridization. J Am Chem Soc 125:8102–8103
27. Demers LM, Mirkin CA, Mucic RC, Reynolds RA 3rd, Letsinger RL, Elghanian R, Viswanadham G (2000) A fluorescence-based method for determining the surface coverage and hybridization efficiency of thiol-capped oligonucleotides bound to gold thin films and nanoparticles. Anal Chem 72:5535–5541
28. Bayley H (2009) Membrane-protein structure: piercing insights. Nature 459:651–652
29. Garaj S, Hubbard W, Reina A, Kong J, Branton D, Golovchenko J (2010) Graphene as a subnanometer trans-electrode membrane. Nature 467:190–193
30. Storm AJ, Chen JH, Ling XS, Zandbergen HW, Dekker C (2003) Fabrication of solid-state nanopores with single-nanometre precision. Nat Mater 2:537–540
31. Stroeve P, Ileri N (2011) Biotechnical and other applications of nanoporous membranes. Trends Biotechnol 29:259–266
32. Capp MW, Pegram LM, Saecker RM, Kratz M, Riccardi D, Wendorff T, Cannon JG, Record MT Jr (2009) Interactions of the osmolyte glycine betaine with molecular surfaces in water: thermodynamics, structural interpretation, and prediction of m-values. Biochemistry 48:10372–10379
33. Knowles DB, LaCroix AS, Deines NF, Shkel I, Record MT Jr (2011) Separation of preferential interaction and excluded volume effects on DNA duplex and hairpin stability. Proc Natl Acad Sci USA 108:12699–12704
34. Hong J, Capp MW, Saecker RM, Record MT Jr (2005) Use of urea and glycine betaine to quantify coupled folding and probe the burial of DNA phosphates in lac repressor-lac operator binding. Biochemistry 44:16896–16911
35. Koumoto K, Ochiai H, Sugimoto N (2008) Structural effect of synthetic zwitterionic cosolutes on the stability of DNA duplexes. Tetrahedron 64:168–174
36. Uversky VN, Li J, Fink AL (2001) Trimethylamine-N-oxide-induced folding of alpha-synuclein. FEBS Lett 509:31–35
37. Miklos AC, Li C, Sharaf NG, Pielak GJ (2010) Volume exclusion and soft interaction effects on protein stability under crowded conditions. Biochemistry 49:6984–6991

38. Spink CH, Chaires JB (1995) Selective stabilization of triplex DNA by poly(ethylene glycols). J Am Chem Soc 117:12887–12888
39. Goobes R, Minsky A (2001) Thermodynamic aspects of triplex DNA formation in crowded environments. J Am Chem Soc 123:12692–12693
40. Goobes R, Kahana N, Cohen O, Minsky A (2003) Metabolic buffering exerted by macromolecular crowding on DNA-DNA interactions: origin and physiological significance. Biochemistry 42:2431–2440
41. Kitts PA, Nash HA (1987) Homology-dependent interactions in phage lambda site-specific recombination. Nature 329:346–348
42. Lu M, Guo Q, Marky LA, Seeman NC, Kallenbach NR (1992) Thermodynamics of DNA branching. J Mol Biol 223:781–789
43. Muhuri S, Mimura K, Miyoshi D, Sugimoto N (2009) Stabilization of three-way junctions of DNA under molecular crowding conditions. J Am Chem Soc 131:9268–9280
44. Shiman R, Draper DE (2000) Stabilization of RNA tertiary structure by monovalent cations. J Mol Biol 302:79–91
45. Semrad K, Green R (2002) Osmolytes stimulate the reconstitution of functional 50S ribosomes from in vitro transcripts of *Escherichia coli* 23S rRNA. RNA 8:401–411
46. Gluick TC, Yadav S (2003) Trimethylamine N-oxide stabilizes RNA tertiary structure and attenuates the denaturating effects of urea. J Am Chem Soc 125:4418–4419
47. Lambert D, Leipply D, Draper DE (2010) The osmolyte TMAO stabilizes native RNA tertiary structures in the absence of Mg^{2+}: evidence for a large barrier to folding from phosphate dehydration. J Mol Biol 404:138–157
48. Pincus DL, Hyeon C, Thirumalai D (2008) Effects of trimethylamine N-oxide (TMAO) and crowding agents on the stability of RNA hairpins. J Am Chem Soc 130:7364–7372
49. Denesyuk NA, Thirumalai D (2011) Crowding promotes the switch from hairpin to pseudoknot conformation in human telomerase RNA. J Am Chem Soc 133:11858–11861
50. Tobe S, Heams T, Vergne J, Herve G, Maurel MC (2005) The catalytic mechanism of hairpin ribozyme studied by hydrostatic pressure. Nucleic Acids Res 33:2557–2564
51. Herve G, Tobe S, Heams T, Vergne J, Maurel MC (2006) Hydrostatic and osmotic pressure study of the hairpin ribozyme. Biochim Biophys Acta 1764:573–577
52. Downey CD, Crisman RL, Randolph TW, Pardi A (2007) Influence of hydrostatic pressure and cosolutes on RNA tertiary structure. J Am Chem Soc 129:9290–9291
53. Kilburn D, Roh JH, Guo L, Briber RM, Woodson SA (2010) Molecular crowding stabilizes folded RNA structure by the excluded volume effect. J Am Chem Soc 132:8690–8696
54. Nakano S, Karimata HT, Kitagawa Y, Sugimoto N (2009) Facilitation of RNA enzyme activity in the molecular crowding media of cosolutes. J Am Chem Soc 131:16881–16888
55. Keniry MA (2000) Quadruplex structures in nucleic acids. Biopolymers 56:123–146
56. Simonsson T (2001) G-quadruplex DNA structures – variations on a theme. Biol Chem 382:621–628
57. Qin Y, Hurley LH (2008) Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. Biochimie 90:1149–1171
58. Dai J, Carver M, Yang D (2008) Polymorphism of human telomeric quadruplex structures. Biochimie 90:1172–1183
59. Lane AN, Chaires JB, Gray RD, Trent JO (2008) Stability and kinetics of G-quadruplex structures. Nucleic Acids Res 36:5482–5515
60. Chen F-M (1992) Sr^{2+} facilitates intermolecular G-quadruplex formation of telomeric sequences. Biochemistry 31:3769–3776
61. Wang Y, Patel DJ (1992) Guanine residues in $d(T_2AG_3)$ and $d(T_2G_4)$ form parallel-stranded potassium cation stabilized G-quadruplexes with anti glycosidic torsion angles in solution. Biochemistry 31:8112–8119
62. Schultze P, Hud NV, Smith FW, Feigon J (1999) The effect of sodium, potassium and ammonium ions on the conformation of the dimeric quadruplex formed by the *Oxytricha nova* telomere repeat oligonucleotide $d(G_4T_4G_4)$. Nucleic Acids Res 27:3018–3028

63. Kankia BI, Marky LA (2001) Folding of the thrombin aptamer into a G-quadruplex with Sr^{2+}: stability, heat, and hydration. J Am Chem Soc 123:10799–10804
64. Miyoshi D, Nakao A, Sugimoto N (2003) Structural transition from antiparallel to parallel G-quadruplex of $d(G_4T_4G_4)$ induced by Ca^{2+}. Nucleic Acids Res 31:1156–1163
65. Wu G, Wong A, Gan Z, Davis JT (2003) Direct detection of potassium cations bound to G-quadruplex structures by solid-state 39K NMR at 19.6 T. J Am Chem Soc 125:7182–7183
66. Sket P, Crnugelj M, Plavec J (2004) $d(G_3T_4G_4)$ forms unusual dimeric G-quadruplex structure with the same general fold in the presence of K^+, Na^+ or NH_4^+ ions. Bioorg Med Chem 12:5735–5744
67. Ida R, Wu G (2005) Solid-state 87Rb NMR signatures for rubidium cations bound to a G-quadruplex. Chem Commun 4294–4296
68. Gill ML, Strobel SA, Loria JP (2005) 205TI NMR methods for the characterization of monovalent cation binding to nucleic acids. J Am Chem Soc 127:16723–16732
69. Gray RD, Chaires JB (2008) Kinetics and mechanism of K^+ and Na^+-induced folding of models of human telomeric DNA into G-quadruplex structures. Nucleic Acids Res 36:4191–4203
70. Wang Y, Patel DJ (1993) Solution structure of the human telomeric repeat $d[AG_3(T_2AG_3)_3]$ G-tetraplex. Structure 1:263–282
71. Parkinson GN, Lee MP, Neidle S (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. Nature 417:876–880
72. Ambrus A, Chen D, Dai J, Bialis T, Jones RA, Yang D (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. Nucleic Acids Res 34:2723–2735
73. Luu KN, Phan AT, Kuryavyi V, Lacroix L, Patel DJ (2006) Structure of the human telomere in K+ solution: an intramolecular (3+1) G-quadruplex scaffold. J Am Chem Soc 128:9963–9970
74. Xu Y, Noguchi Y, Sugiyama H (2006) The new models of the human telomere $d[AGGG(TTAGGG)_3]$ in K^+ solution. Bioorg Med Chem 14:5584–5591
75. Miyoshi D, Nakao A, Sugimoto N (2002) Molecular crowding regulates the structural switch of the DNA G-quadruplex. Biochemistry 41:15017–15024
76. Li J, Correia JJ, Wang L, Trent JO, Chaires JB (2005) Not so crystal clear: the structure of the human telomere G-quadruplex in solution differs from that present in a crystal. Nucleic Acids Res 33:4649–4659
77. Miyoshi D, Karimata H, Sugimoto N (2005) Drastic effect of a single base difference between human and Tetrahymena telomere sequences on their structures under molecular crowding conditions. Angew Chem Int Ed 44:3740–3744
78. Xue Y, Kan ZY, Wang Q, Yao Y, Liu J, Hao YH, Tan Z (2007) Human telomeric DNA forms parallel-stranded intramolecular G-quadruplex in K^+ solution under molecular crowding condition. J Am Chem Soc 129:11185–11191
79. Zhou J, Wei C, Jia G, Wang X, Tang Q, Feng Z, Li C (2008) The structural transition and compaction of human telomeric G-quadruplex induced by excluded volume effect under cation-deficient conditions. Biophys Chem 136:124–127
80. Renciuk D, Kejnovská I, Skoláková P, Bednárová K, Motlová J, Vorlícková M (2009) Arrangements of human telomere DNA quadruplex in physiologically relevant K^+ solutions. Nucleic Acids Res 37:6625–6634
81. Xu L, Feng S, Zhou X (2011) Human telomeric G-quadruplexes undergo dynamic conversion in a molecular crowding environment. Chem Commun 47:3517–3519
82. Heddi B, Phan AT (2011) Structure of human telomeric DNA in crowded solution. J Am Chem Soc 133:9824–9833
83. Hänsel R, Löhr F, Foldynová-Trantírková S, Bamberg E, Trantírek L, Dötsch V (2011) The parallel G-quadruplex structure of vertebrate telomeric repeat sequences is not the preferred folding topology under physiological conditions. Nucleic Acids Res 39:5768–5775
84. Miyoshi D, Karimata H, Sugimoto N (2006) Hydration regulates thermodynamics of G-quadruplex formation under molecular crowding conditions. J Am Chem Soc 128:7957–7963

85. Miller MC, Buscaglia R, Chaires JB, Lane AN, Trent JO (2010) Hydration is a major determinant of the G-quadruplex stability and conformation of the human telomere 3′ Sequence of $d(AG_3(TTAG_3)_3)$. J Am Chem Soc 132:17105–17107
86. Phan AT, Mergny JL (2002) Human telomeric DNA: G-quadruplex, i-motif and Watson–Crick double helix. Nucleic Acids Res 30:4618–4625
87. Miyoshi D, Matsumura S, Nakano S, Sugimoto N (2004) Duplex dissociation of telomere DNAs induced by molecular crowding. J Am Chem Soc 126:165–169
88. Kan Z-Y, Yao Y, Wang P, Li X-H, Hao Y-H, Tan Z (2006) Molecular crowding induces telomere G-quadruplex formation under salt-deficient conditions and enhances its competition with duplex formation. Angew Chem Int Ed 45:1629–1632
89. Kan ZY, Lin Y, Wang F, Zhuang XY, Zhao Y, Pang DW, Hao YH, Tan Z (2007) G-quadruplex formation in human telomeric $(TTAGGG)_4$ sequence with complementary strand in close vicinity under molecularly crowded condition. Nucleic Acids Res 35:3646–3653
90. Zhou J, Wei C, Jia G, Wang X, Feng Z, Li C (2009) Human telomeric G-quadruplex formed from duplex under near physiological conditions: spectroscopic evidence and kinetics. Biochimie 91:1104–1111
91. Kumar N, Maiti S (2005) The effect of osmolytes and small molecule on Quadruplex-WC duplex equilibrium: a fluorescence resonance energy transfer study. Nucleic Acids Res 33:6723–6732
92. Zheng KW, Chen Z, Hao YH, Tan Z (2010) Molecular crowding creates an essential environment for the formation of stable G-quadruplexes in long double-stranded DNA. Nucleic Acids Res 38:327–338
93. Li W, Miyoshi D, Nakano S, Sugimoto N (2003) Structural competition involving G-quadruplex DNA and its complement. Biochemistry 42:11736–11744
94. Miyoshi D, Matsumura S, Li W, Sugimoto N (2003) Structural polymorphism of telomeric DNA regulated by pH and divalent cation. Nucleosides Nucleotides Nucleic Acids 22:203–221
95. Kumar N, Maiti S (2004) Quadruplex to Watson–Crick duplex transition of the thrombin binding aptamer: a fluorescence resonance energy transfer study. Biochem Biophys Res Commun 319:759–767
96. Miyoshi D, Inoue M, Sugimoto N (2006) DNA logic gates based on structural polymorphism of telomere DNA molecules responding to chemical input signals. Angew Chem Int Ed 45:7716–7719
97. Arora A, Maiti S (2009) Stability and molecular recognition of quadruplexes with different loop length in the absence and presence of molecular crowding agents. J Phys Chem B 113:8784–8792
98. Zhang DH, Fujimoto T, Saxena S, Yu HQ, Miyoshi D, Sugimoto N (2010) Monomorphic RNA G-quadruplex and polymorphic DNA G-quadruplex structures responding to cellular environmental factors. Biochemistry 49:4554–4563
99. Miyoshi D, Nakamura K, Tateishi-Karimata H, Ohmichi T, Sugimoto N (2009) Hydration of Watson–Crick base pairs and dehydration of Hoogsteen base pairs inducing structural polymorphism under molecular crowding conditions. J Am Chem Soc 131:3522–3531
100. Pramanik S, Nakamura K, Usui K, Nakano S, Saxena S, Matsui J, Miyoshi D, Sugimoto N (2011) Thermodynamic stability of Hoogsteen and Watson–Crick base pairs in the presence of histone H3-mimicking peptide. Chem Commun 47:2790–2792
101. Bloomfield VA, Crothers DM, Tinoco I Jr (2000) Nucleic acids structures, properties, and functions. University Science Books, Sausalito, California
102. Foley PL, Wilson DB, Shuler ML (2010) Macromolecular crowding can account for RNase-sensitive constraint of bacterial nucleoid structure. Biochem Biophys Res Commun 395:42–47
103. Olsen CM, Gmeiner WH, Marky LA (2006) Unfolding of G-quadruplexes: energetic, and ion and water contributions of G-quartet stacking. J Phys Chem B 110:6962–6969

104. Olsen CM, Lee HT, Marky LA (2009) Unfolding thermodynamics of intramolecular G-quadruplexes: base sequence contributions of the loops. J Phys Chem B 113:2587–2595
105. Arora A, Maiti S (2009) Differential biophysical behavior of human telomeric RNA and DNA quadruplex. J Phys Chem B 113:10515–10520
106. Zimmerman SB, Trach SO (1988) Macromolecular crowding extends the range of conditions under which DNA polymerase is functional. Biochim Biophys Acta 949:297–304
107. Jarvis TC, Ring DM, Daube SS, von Hippel PH (1990) "Macromolecular crowding": thermodynamic consequences for protein–protein interactions within the T4 DNA replication complex. J Biol Chem 265:15160–15167
108. Sasaki Y, Miyoshi D, Sugimoto N (2006) Effect of molecular crowding on DNA polymerase activity. Biotechnol J 1:440–446
109. Sasaki Y, Miyoshi D, Sugimoto N (2007) Regulation of DNA nucleases by molecular crowding. Nucleic Acids Res 35:4086–4093
110. Yu HQ, Zhang DH, Gu XB, Miyoshi D, Sugimoto N (2008) Regulation of telomerase activity by the thermodynamic stability of a DNA x RNA hybrid. Angew Chem Int Ed 47:9034–9038
111. Chen Z, Zheng KW, Hao YH, Tan Z (2009) Reduced or diminished stabilization of the telomere G-quadruplex and inhibition of telomerase by small chemical ligands under molecular crowding condition. J Am Chem Soc 131:10430–10438
112. Martino L, Pagano B, Fotticchia I, Neidle S, Giancola C (2009) Shedding light on the interaction between TMPyP4 and human telomeric quadruplexes. J Phys Chem B 113:14779–14786
113. Wei C, Wang J, Zhang M (2010) Spectroscopic study on the binding of porphyrins to $(G_4T_4G_4)_4$ parallel G-quadruplex. Biophys Chem 148:51–55
114. Wei C, Jia G, Zhou J, Han G, Li C (2009) Evidence for the binding mode of porphyrins to G-quadruplex DNA. Phys Chem Chem Phys 11:4025–4032
115. Li W, Zhang M, Zhang JL, Li HQ, Zhang XC, Sun Q, Qiu CM (2006) Interactions of daidzin with intramolecular G-quadruplex. FEBS Lett 580:4905–4910
116. Petraccone L, Fotticchia I, Cummaro A, Pagano B, Ginnari-Satriani L, Haider S, Randazzo A, Novellino E, Neidle S, Giancola C (2011) The triazatruxene derivative azatrux binds to the parallel form of the human telomeric G-quadruplex under molecular crowding conditions: biophysical and molecular modeling studies. Biochimie 93:1318–1327
117. Dominak LM, Omiatek DM, Gundermann EL, Heien ML, Keating CD (2010) Polymeric crowding agents improve passive biomacromolecule encapsulation in lipid vesicles. Langmuir 26:13195–13200
118. Zhao C, Ren J, Qu X (2008) Single-walled carbon nanotubes binding to human telomeric i-motif DNA under molecular-crowding conditions: more water molecules released. Chemistry 14:5435–5439
119. Khripin CY, Arnold-Medabalimi N, Zheng M (2011) Molecular-crowding-induced clustering of DNA-wrapped carbon nanotubes for facile length fractionation. ACS Nano 5:8258–8266
120. Goodrich GP, Helfrich MR, Overberg JJ, Keating CD (2004) Effect of macromolecular crowding on DNA:Au nanoparticle bioconjugate assembly. Langmuir 20:10246–10251
121. Zaki A, Dave N, Liu J (2012) Amplifying the macromolecular crowding effect using nanoparticles. J Am Chem Soc 134:35–38
122. Ricci F, Lai RY, Heeger AJ, Plaxco KW, Sumner JJ (2007) Effect of molecular crowding on the response of an electrochemical DNA sensor. Langmuir 23:6827–6834
123. Feng B, Frykholm K, Nordén B, Westerlund F (2010) DNA strand exchange catalyzed by molecular crowding in PEG solutions. Chem Commun 46:8231–8233
124. Zhang C, Shao PG, van Kan JA, van der Maarel JR (2009) Macromolecular crowding induced elongation and compaction of single DNA molecules confined in a nanochannel. Proc Natl Acad Sci USA 106:16651–16666

Top Curr Chem (2013) 330: 111–178
DOI: 10.1007/128_2012_346

Published online: 12 August 2012

Visualizing the Quadruplex: From Fluorescent Ligands to Light-Up Probes

Eric Largy, Anton Granzhan, Florian Hamon, Daniela Verga, and Marie-Paule Teulade-Fichou

Abstract Detection of quadruplex structures by visual methods is a major challenge of the quadruplex nucleic acid research area. Consequently, considerable efforts are under way for the discovery of quadruplex specific agents endowed with fluorescence properties. In this review chapter we propose a comprehensive and critical overview of the diverse molecular design and strategies that have been described to identify quadruplex-selective fluorescent probes. Innovative compounds as well as classical DNA dyes are reviewed. The compounds have been divided into three classes: (1) "light-up" probes that display a strong enhancement upon G4 binding, (2) "light-off" probes that display a decreased fluorescence upon binding, and (3) permanent probes ("tagged" G4-binders) that exhibit no variation of fluorescence but display quadruplex binding specificity. The labeling performances of probes in various analytical contexts (in solution, in gel, at the level of chromosomes, and in fixed cells) are also reported and commented on when available. Finally we address the strengths and weaknesses of each probe class and highlight the critical features that must be addressed in developing a practicable quadruplex-specific labeling agent.

Keywords Fluorescence · Fluorescent probes · Nucleic acids · Quadruplex DNA

Contents

1 Introduction 112
2 "Light-Up" Probes 115
2.1 Classical DNA Probes 115
2.2 Triphenylmethane Dyes 128
2.3 Carbazoles 130

E. Largy, A. Granzhan, F. Hamon, D. Verga and M.-P. Teulade-Fichou (✉)
UMR 176 CNRS "Synthèse, Conception et Vectorisation de Biomolécules", Institut Curie, Centre de Recherche, 91405 Orsay, France
e-mail: mp.teulade-fichou@curie.fr

2.4 Porphyrins 134
2.5 Probes Operating via Aggregation-Induced Emission 137
2.6 Miscellaneous 140
2.7 Metal Complexes 145
3 "Light-Off" Probes 158
3.1 Heterocyclic Ligands with a Neutral Core 160
3.2 Heterocyclic Ligands with a Cationic Core 162
3.3 Perylene Diimide Derivatives 164
3.4 Anthracyclines 164
3.5 Classical Organic Dyes 165
3.6 Metal Complexes 166
4 Permanent ("Tagged") Quadruplex-DNA Probes 167
5 Conclusions and Perspectives 170
References 171

Abbreviations

AIE	Aggregation-induced emission
CD	Circular dichroism
ct DNA	Calf thymus DNA
ds DNA	Double-stranded DNA
ESIPT	Excited-state intramolecular proton transfer
FID	Fluorescent intercalator displacement
FRET	Förster resonance energy transfer
G4-DNA	Quadruplex DNA
ITC	Isothermal titration calorimetry
PAGE	Polyacrylamide gel electrophoresis
PNA	Peptide nucleic acid
ss DNA	Single-stranded DNA
SPR	Surface plasmon resonance
TBA	Thrombin-binding aptamer
TO	Thiazole orange
TPM	Triphenylmethane

1 Introduction

Single-stranded regions of nucleic acid containing repeats of bases are known to form alternative secondary structures due to unusual (i.e., non-Watson–Crick) base pairing. These alternative structures are believed to act as roadblocks to proteins processing single-stranded DNA templates transiently formed during replication, homologous recombination, repair, and transcription. In the same manner, transcriptional machineries that use single-stranded messenger RNA as substrate for translation and splicing may also be affected. However, although alternative structures are currently recognized to interfere with the transfer of genetic

information, the question as to whether they participate in a natural regulation mechanism or are simply toxic elements is still open [1–7].

Currently the most well-studied and characterized non-B-DNA structures are the four-stranded assemblies called G-quadruplexes (abbreviated G4), which form by fold-over of single-stranded sequences containing guanines repeats and are the subject of this volume. G-Quadruplexes are composed of π-stacks of guanine quartets held together by H-bonding and by monovalent metal cations (K^+ in a cellular context) sandwiched between the quartets (for structural details see reviews [8–11]).

Structural recognition of quadruplex nucleic acids by small molecules has been the focus of tremendous developments since the pioneering work of S. Neidle, L. Hurley et al. who paved the way by targeting quadruplex DNA with anthraquinone derivatives [12]. Since then this research area is progressing rapidly in the desire to find small molecules able to probe the formation of quadruplexes in a biological context and to help in deciphering their biological functions. Another objective of this field is to discover new targeted pharmacological agents able to act in a sequence/structure specific manner on DNA or RNA with resulting anticancer properties [13–15].

Although well-studied in vitro, evidence of the formation of quadruplexes in vivo remains mostly indirect but is currently supported by robust biochemical and molecular genetics data [1–7, 16]. Logically, the question of direct visualization of quadruplexes arises. Compelling evidence has been provided by observation of G4-DNA in human DNA by electron and atomic force microscopy [17–20] and by immunochemistry in ciliates [21]. In parallel, the search for fluorescent compounds that could signal the presence of a quadruplex structure is a matter of great interest, especially in view of the increasing use of fluorescent probes in cellular biology and the emergence of new fluorescence-based microscopies [22, 23]. However, currently, direct visualization of quadruplexes in a functional biological system is a challenging task accompanied with technical difficulties and conceptual uncertainties. Indeed, one should bear in mind that only large domains with a high density of quadruplexes able to form fluorescent foci could be imaged by microscopy, since detection of a single monomeric quadruplex unit is largely below the actual resolution; additionally, whether such domains do form and under what conditions is poorly understood. Nonetheless, the first absolute requirement of this quest is the availability of highly specific quadruplex-binding fluorescent dyes. Consequently, the chemistry of fluorescent compounds able to target quadruplexes experienced extensive development during the past decade.

In this review we describe the considerable work in terms of molecular design that has produced a large variety of quadruplex-interactive agents able to achieve G4-DNA labeling in different analytical contexts (e.g., from solution to gel electrophoresis and cellular imaging) and with various efficiencies. The compounds have been divided into three classes: (1) “light-up” probes, i.e., compounds that display a strongly enhanced fluorescence upon binding, (2) “light-off” probes, i.e., compounds that display a strongly decreased fluorescence upon binding, and (3) permanent probes, i.e., compounds the fluorescence of which is not modified by binding, but exhibiting exquisite specificity; these are mainly fluorescently tagged

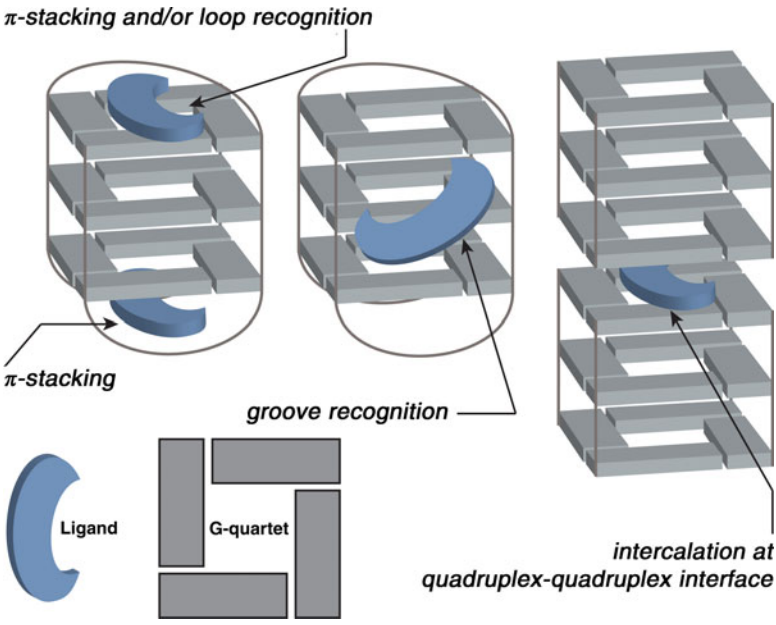


Fig. 1 Binding modes of ligands to quadruplex DNA

quadruplex binders. Finally, we address the strengths and weaknesses of each probe class and examine the issues and highlight the critical probe design features that must be addressed in developing a practicable G4-DNA probe (especially as a function of the intended application).

In order to help understand the diverse molecular designs of the probes described below, it is necessary to recall that quadruplexes may accommodate small molecules in several ways as depicted in Fig. 1. π-Stacking on external G-quartets on the top or bottom of a quadruplex unit is the most well-known mode characterized both in solid state and in solution [24–26]. In domains containing repeats of quadruplex units, sandwiching of a ligand between two units is likely to occur, as shown by X-ray studies [27, 28] and suggested by indirect evidence [20]. The stacking mode that may eventually involve interactions with surrounding loops is mainly the feature of planar rigid aromatics. Alternatively, insertion in the grooves of quadruplexes, that possess four grooves, is to be considered, especially for nonplanar, flexible molecules as shown by NMR [29] and CD in recent studies [30, 31]. Finally, quadruplexes provide a polyanionic scaffold that is capable of inducing external binding of cationic molecules in more or less ordered aggregates as is known for duplex DNA and certain dyes [32]. We will see that the design of fluorescent G4-DNA ligands relies on these different binding sites that provide various environments differing by their polarity and their capacity to immobilize external ligands and to isolate them from interaction with water – three parameters that strongly affect the fluorescence of dyes.

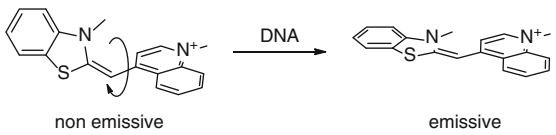


Fig. 2 Binding to DNA immobilizes cyanine dyes (here thiazole orange) and hinders the non-radiative deactivation pathways through restriction of the rotation around the conjugated bond, leading to enhanced fluorescence

2 "Light-Up" Probes

2.1 Classical DNA Probes

2.1.1 Cyanine Dyes

The cyanine (polymethine) dyes, which are among the oldest synthetic dyes, have been widely used as fluorescent probes for nucleic acids as they may display a large fluorescent enhancement upon binding. Their binding modes may be complex depending strongly on the length of the polymethine bridge and on the nucleic acid conformation. Thus, intercalation, dimerization in minor grooves, and template-directed external self-stacking have been characterized [33, 34]. Asymmetric cyanines are built from two distinct heteroaromatic systems connected through a conjugated monomethine or polymethine bridge. Short-bridged cyanines (monomethine and trimethine dyes) are often characterized by a turn-on fluorescence upon binding to DNA. This phenomenon is rationalized by the restriction of the rotation of this conjugated bridge, thus preventing nonradiative relaxation processes from taking place (Fig. 2) [35, 36].

Thiazole Orange

Thiazole orange (TO, Fig. 2) is a well-known monomethine dye that binds to almost all forms of nucleic acids. Therefore TO has been used in a variety of applications such as RNA staining [37], agarose gel DNA staining [38], and capillary electrophoresis [39]. TO represents a paradigm of nucleic acid light-up probes since its fluorescence quantum yield is about 2×10^{-4} when the dye is free in solution but increases to around 0.1 when bound to ds DNA and poly(dA), or 0.4 in the case of poly(dG) [35, 40]. Thus, TO exhibits a fluorescence enhancement upon binding, which varies from 50-fold to 1,000-fold depending on the sequence and conformation of the nucleic acid.

TO was firstly used as a fluorescent indicator for quadruplex DNA in the G4 fluorescent intercalator-displacement (FID) assay [41], by analogy with the previous work of Boger et al. on DNA hairpins [42, 43]. It was found that TO becomes highly fluorescent (500–3,000-fold increase, $\lambda_{ex} = 501$ nm, $\lambda_{max} = 539$ nm; Fig. 3)

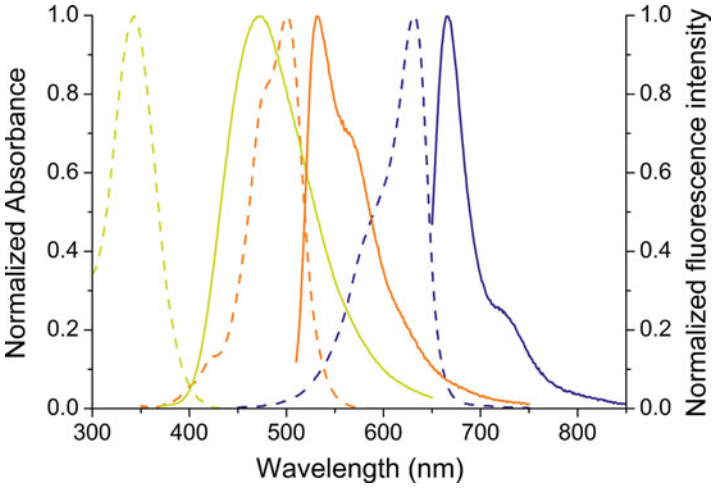


Fig. 3 Normalized absorbance (*dashed lines*) and fluorescence spectra (*solid lines*) of Hoechst 33258 (*yellow*), thiazole orange (*orange*), and TO-PRO-3 (*blue*) in the presence of c-kit2 (concentration of dye =0.5 μM; dye/DNA ratio =2). Reprinted with permission from [44]. Copyright 2011 Springer Science+Business Media

when bound to various intramolecular quadruplex structures including those formed from the human telomeric repeats, oncogene promoters, and minisatellite sequences (c-myc, c-kit, CEB1) [44–46].[1] Significantly lower fluorescent enhancements (around 60–70-fold) were found with tetramolecular parallel quadruplexes (TG_3T, TG_4T, and TG_5T) although these structures also seem to accommodate TO, as estimated by ESI-MS measurements [44, 46]. Apparent affinities of TO to quadruplex and duplex DNA are close, with K_d values falling in the micromolar range. Fluorescence titrations, Job plot analysis, and mass spectrometry supported a 1:1 stoichiometry model between TO and several intramolecular quadruplex DNAs and 2:1 with tetramolecular structures TG_3T, TG_4T, and TG_5T. The predominant binding mode is likely to be π-stacking on external G-tetrads, and in this regard the lower fluorescence of TO bound to parallel quadruplexes would be explained by the absence of loops resulting in a poorer immobilization of the dye and/or poorer isolation from water. However, groove binding or external stacking cannot be completely excluded in these cases. We will see below that other light-up probes also show lower fluorescence enhancement in tetramolecular quadruplexes.

Finally, TO was found to label pretty well RNA quadruplexes and dimeric quadruplex matrices made of longer DNA or RNA sequences ([47] and unpublished data). The displacement of TO from G4-DNA has been exploited to implement a

[1] Experiments performed in lithium cacodylate buffer, at pH 7.2 with 100 mM KCl or NaCl.

high-throughput screening assay for the discovery of new G4 binders via library screening [48].

In particular conditions (pH 8.0 in the absence of monovalent cations), TO seems to interact and to fluoresce preferentially with G-wires (i.e., long tetramolecular quadruplexes) than with double helical DNA [49]. In these conditions, TO and G-wires form stable complexes in gel experiments whereas this is not the case with ds DNA.

TO has been used in a quadruplex-templated FRET experiment to probe quadruplex conformations through a ternary interaction system [50]. In this system, TO and a quadruplex-specific binder (BOQ_1 or MMQ_1; cf. Sect. 3.1), displaying appropriate photophysical properties (i.e., emission around 450–500 nm, overlapping the absorption of TO) form a donor–acceptor pair, with TO being the acceptor. Quadruplex structures have two external quartets and can thus accommodate both compounds close enough (about 10 Å) for FRET to occur. When the human telomeric quadruplex structure is added to a mixture of BOQ_1 (or MMQ_1) and TO, a strong decrease of BOQ_1 (MMQ_1) fluorescence is observed, while TO emission is greatly increased (see footnote 1). Importantly, the FRET response was found to be specific for the quadruplex structure, indicating that the quadruplex matrix provides a more efficient positioning of the two FRET partners as compared to a duplex matrix.

In vitro fluorescent detection of 5′,3′-cyclic diguanylic acid (c-di-GMP) using TO has recently been achieved [51]. While TO is not fluorescent in the presence of simple nucleotides (GTP, GMP, cGMP), it induces the formation of quadruplex structures composed of c-di-GMP G-tetrads via intercalation or external stacking. Higher cation concentration (up to 1 M), and Na^+ preferentially to K^+ and NH_4^+, were shown to increase the TO fluorescence since it promotes c-di-GMP aggregation. In conclusion, the exceptional fluorescence properties of TO and its strong binding to many quadruplexes make it a very efficient quadruplex stain for in vitro experiments, in particular for fluorescence-based screening assays.

In cells, TO was found to stain preferentially nucleoli (but not nuclear DNA) as well as the cytoplasm [49]. This is consistent with the fact that TO and derivatives (SYTO dyes) are known to stain preferentially RNAs and are commercialized on this purpose [52]. For this reason, and owing to its absence of significant binding selectivity for biological quadruplexes it appears rather irrelevant to use this dye for probing quadruplexes inside cells, although the presence of quadruplex RNA (and also DNA) in the nucleoli is not excluded.

TO Derivatives

In order to probe specifically quadruplex DNA, TO has been appended to bisquinolinium compounds, ones of the best and most selective quadruplex binders of the pyridodicarboxamide (PDC) series [53–55]. Two designs were conceived: the first, PDC-*L*-TO, consists in linking both components by a flexible linker, whereas the second, PDC-*M*-TO, is built by merging the *N*-methylquinolinium moiety present in both scaffolds (Fig. 4).

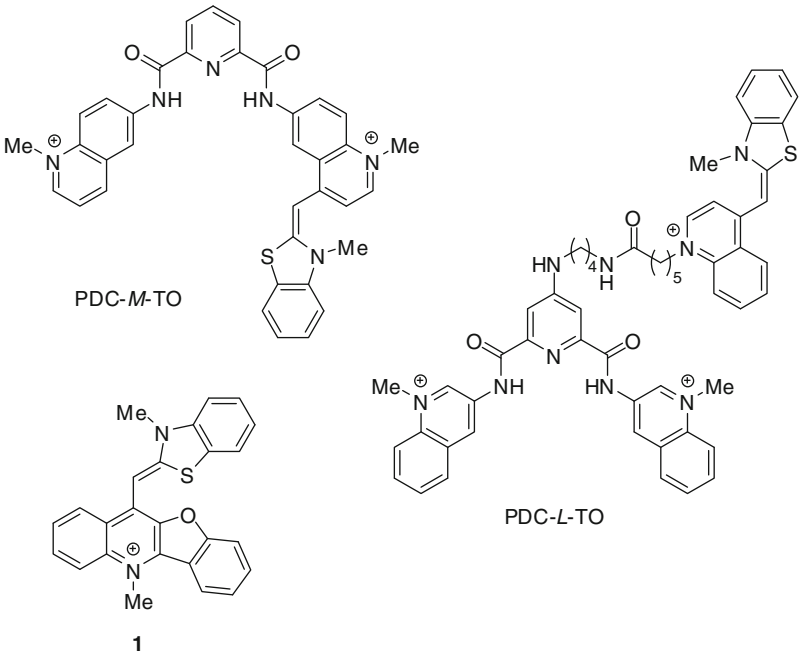


Fig. 4 TO derivatives designed for targeting quadruplex DNA

Both designs afforded fluorescent off/on probes which were different in their selectivity. Thus, PDC-*L*-TO is devoid of selectivity for quadruplex DNA and displays an intense orange–yellow fluorescence ($\lambda_{em} = 530$ nm) both with duplex and quadruplex DNA. The absence of selectivity is attributable to the affinity of TO for duplex DNA that counterbalances the quadruplex preference of the PDC motif. On the other hand, PDC-*M*-TO stabilizes strongly the telomeric quadruplex DNA even in the presence of a strong excess of ds DNA ($\Delta T_{1/2} = 17$ °C). The fluorescence enhancement of PDC-*M*-TO ($\lambda_{ex} = 500$ nm, $\lambda_{em} = 550$ nm) when bound to the telomeric quadruplex (22AG: d[$G_3(T_2AG_3)_3$]) is decreased by about 36% as compared to TO alone, but displays an eightfold selectivity compared to the complex with ds DNA.[2] On this basis, selective staining of quadruplex DNA could be observed in gel electrophoresis, but the sensitivity of the merged compound remained weaker than that of the linked PDC-*L*-TO compound (Fig. 5). This result underlines the difficulties in fluorescence engineering, as the quadruplex selectivity is retained at the expense of the fluorescence quantum yield. In addition, both dyes show some degree of aggregation at high ionic strength. Still, the new merged design described in this study represents an innovative and solid basis for the design of novel G-quadruplex probes.

[2] Experiments performed at pH 7.2 with 1 mM KCl.

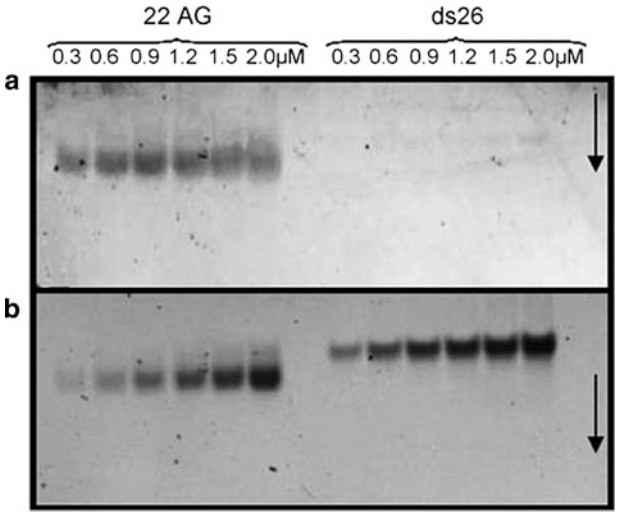


Fig. 5 PAGE of human telomeric quadruplex (22AG) and duplex oligonucleotide (ds26) DNA (0.3–2.0 μM) stained by PDC-*M*-TO (**a**, 5 μM) and post-stained with PDC-*L*-TO (**b**, 2 μM) (λ_{ex} = 532 nm, λ_{max} = 550 nm). Reprinted with permission from [53]. Copyright 2009 John Wiley and Sons

Based on the same concept, another group designed a benzothiazole-substituted benzofuroquinolinium dye **1** (Fig. 4) [56], in which TO has been merged with benzofuroquinoline, a planar aromatic system known for its quadruplex-interactive properties [57]. Both absorption and fluorescence of this new hybrid derivative are significantly shifted towards the red as compared to TO (λ_{ex} = 527 nm, λ_{max} = 573 nm), thereby indicating a higher degree of conjugation. As expected, the dye is poorly fluorescent in water (Φ_F = 0.0012) and exhibits a bright orange fluorescence when bound to quadruplex DNA (λ_{ex} = 527 nm, λ_{max} = 573 nm, Φ_F = 0.25 and 0.18 for the telomeric repeat and c-myc, respectively). Conversely, much smaller changes were observed with duplex DNA (26 bp oligonucleotide and calf thymus DNA) and none with the bovine serum albumin protein (Fig. 6). Moreover, quadruplex and duplex DNA could be easily distinguished in gel electrophoresis using this hybrid molecule as a stain.[3]

The dye is supposed to bind by π−π end-stacking in a quasi-planar conformation, the rotation about the methine bridge being efficiently restricted. The absorption spectra in the presence and the absence of DNA show the typical blue-shift observed for cyanine aggregates, thereby suggesting that the free dye also aggregates to a certain extent, although this was not considered. Finally, the dye is slightly fluorescent (by a factor of 3) in the presence of single-stranded and double-stranded RNA as compared to G-quadruplex RNA and DNA. Imaging in

[3] Experiments performed in 10 mM Tris–HCl buffer with 60 mM KCl.

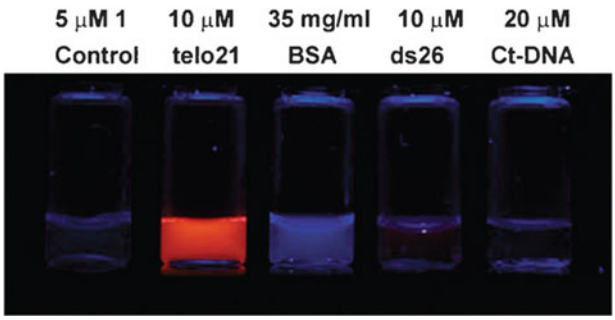


Fig. 6 Pictures of the benzofuroquinolinium dye (5 μM) in the presence of human telomeric quadruplex (telo21), BSA protein, duplex DNA oligonucleotide (ds26) and calf thymus DNA (ct DNA) upon irradiation at 302 nm. Reprinted from [56] with permission of The Royal Society of Chemistry

fixed cells (MCF7) indicates that the dye is localized in the nuclei with faint staining of genomic chromatin.

Other Cyanine Dyes

SYBR Green I is an asymmetric cyanine dye, structurally and photophysically similar to TO ($\lambda_{ex} = 497$ nm, $\lambda_{max} = 520$ nm upon binding to duplex DNA), and widely used for "in-gel" staining of duplex and, to a lower extent, single-stranded DNA (Fig. 7) [58].

SYBR Green I has recently been used in a quadruplex and K^+ detection system in cooperation with a water-soluble cationic conjugated polymer [59]. The polymer allows a sixfold fluorescence amplification of the dye bound to quadruplex through a FRET effect (Fig. 8),[4] but not with single stranded DNA (in the absence of potassium). However, the same effect has been observed previously in an independent study using duplex DNA [60].

TO-PRO-3 is a trimethine analogue of TO, known to bind strongly duplex DNA with a high fluorescence enhancement (Fig. 7) [61]. TO-PRO-3 has recently been used as an alternative to TO in the G4-FID method to evaluate the putative G4-DNA binders absorbing around 500 nm, since its absorbance and fluorescence spectra are red-shifted as compared to TO ($\lambda_{ex} = 642$ nm, $\lambda_{max} = 718$ nm upon binding to the c-kit2 quadruplex; Fig. 3) [44]. Like TO, it binds nucleic acids with micromolar K_d, without strong discrimination between various quadruplex and duplex structures but with high fluorescence enhancements, ranging from 120 to 250 depending on the DNA sequence and structure. The binding modes of TO-PRO-3 to duplex DNA were shown to be multiple, involving both intercalation

[4] Experiments performed in ultra-pure water with 50 mM KCl.

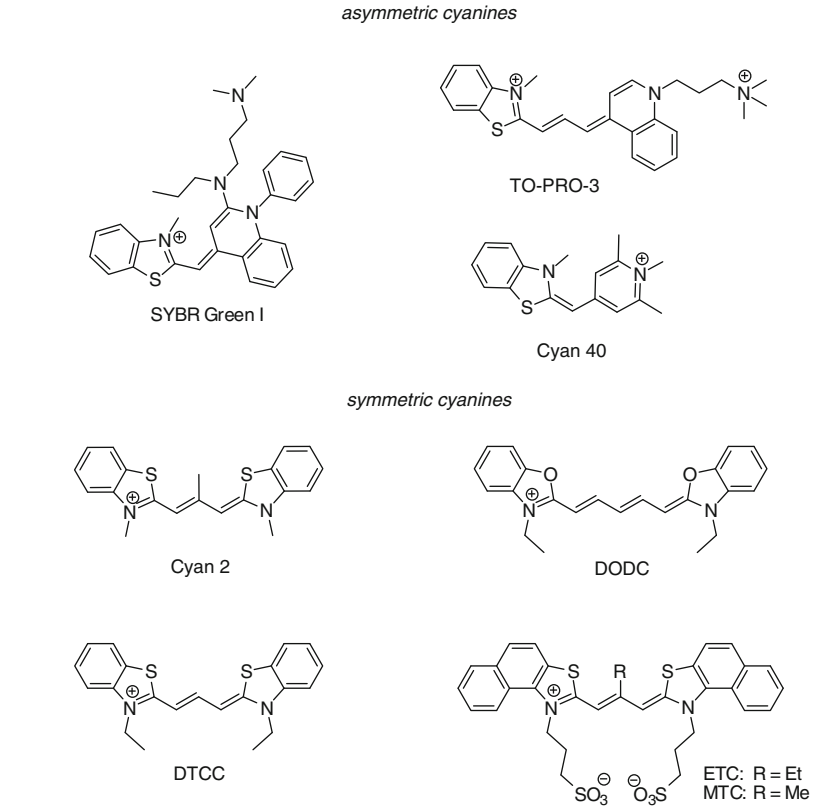


Fig. 7 Cyanine dyes evaluated for recognition of quadruplex DNA

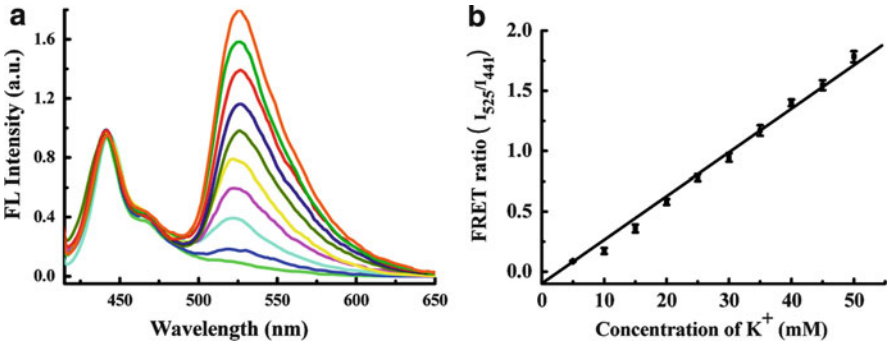


Fig. 8 (**a**) Emission spectra and (**b**) FRET ratio (F_{525}/F_{441}) of SYBR Green I (acceptor) and cationic conjugated polymer (donor) with a quadruplex-forming sequence in the presence of increasing concentrations of K^+ (0–50 mM). Reprinted with permission from [59]. Copyright 2010 American Chemical Society

and groove dimerization/aggregation [34]. Likewise, the binding to quadruplexes might also follow multiple modes, although π-stacking on a G-quartet is most likely.

The monomethine dye Cyan 40 and the trimethine Cyan 2 were examined with respect to their binding to G-wires containing 750 guanine tetrads in the absence of monovalent cations and at pH 8.0 [62]. The fluorescence enhancement of Cyan 2 increases more in the presence of quadruplex (15-fold) and triplex structures than in the presence of duplex DNA. However, it should be noted that both Cyan 2 and Cyan 40 were previously described to act as efficient "light-up" probes for ds DNA [63, 64] and were found to localize in the nucleoli of mammalian cells [25]. Again, this addresses the question of rationalizing cellular localization of G4-DNA stains on the ground of their interaction and selectivity for G-wires that are artificial polymeric quadruplexes without entrapped cations, and thus have no biological relevance.

DODC is a symmetric cyanine that binds to duplex DNA in minor grooves, mainly as a dimer. Its binding to quadruplex DNA was first examined 15 years ago with various dimeric hairpin quadruplex structures (i.e., displaying two diagonal loops) [65]. DODC binds to these structures with micromolar K_d and displays an induced circular dichroism (ICD) signal which may result from groove interactions [30]. More importantly, fluorescence intensity of DODC upon binding was dramatically reduced (about 50%) together with a 6 nm red-shift, but increased (18%) with the tetramolecular parallel TG_4T quadruplex and not significantly changed with other structures (ss DNA, ds DNA, hairpins).[5]

More recently, DODC was used as a fluorescent reporter for quadruplex DNA in a "mix and measure" fluorescence screening assay [66]. A variety of quadruplex structures were examined and, unlike in previous results, DODC was found to exhibit a high fluorescence enhancement (220–500-fold with various concentrations of DNA; $\lambda_{ex} = 579$ nm, $\lambda_{max} = 610$ nm).[6] DTCC, which also binds to minor grooves of ds DNA, was also included in the aforementioned test. Higher fluorescence enhancement were monitored (240–980-fold, $\lambda_{ex} = 554$ nm, $\lambda_{max} = 577$ nm) but with greater amounts of DNA.

ETC (Fig. 7) is an extended aromatic cyanine dye which was found to recognize human telomeric quadruplexes through end-stacking interactions [67, 68]. In aqueous buffer it readily forms J-aggregates ($\lambda_{abs} = 660$ nm) that dissociate in monomer ($\lambda_{abs} = 584$ nm) upon binding to quadruplex DNA.[7] Addition of various quadruplex structures (human telomeric repeat, c-kit1, c-myc, bcl-2) induced a strong fluorescence enhancement of ETC monomer (70-fold, $\lambda_{max} = 600$ nm), and allowed specific detection in gel electrophoresis (PAGE) experiments.[8] Detection at higher (physiological) cation concentration[9] was achieved with MTC, a

[5] Experiments performed in 10 mM Tris buffer with 150 mM NaCl.

[6] Experiments performed in 2.5 mM phosphate buffer, 5 mM Na^+, and 100 mM K^+.

[7] Experiments performed in 10 mM PBS K^+ or Na^+.

[8] Experiments performed with 17.2 mM K^+ and no Na^+.

[9] Experiments performed with 140 mM K^+, 10 mM Na^+, and 2 mM Mg^{2+}.

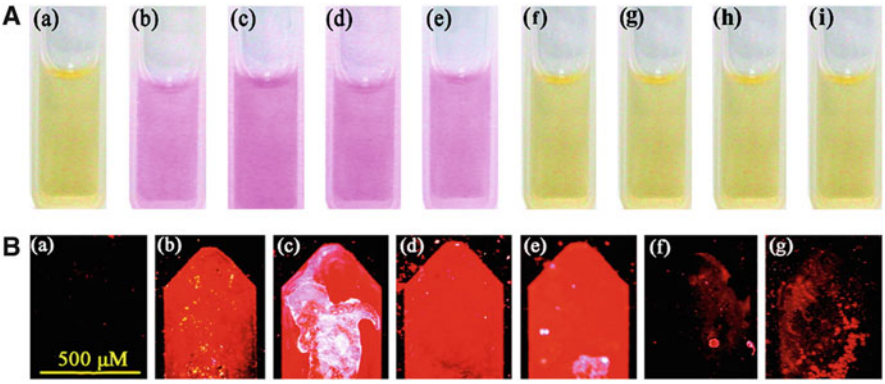


Fig. 9 (**A**) Apparent color of MTC (4 μM) alone (a) and with 20 μM of human telomeric (b), c-myc 2345 (c), c-kit1 (d), bcl-2 (e) quadruplexes, 22-mer oligonucleotide duplex (f), 17-mer single strand (g), calf-thymus (h), and salmon sperm (i) DNA. (**B**) Fluorescence microscopy images of MTC-stained oligonucleotides on Au film (same legend). Adapted with permission from [69]. Copyright 2010 American Chemical Society

close analog of ETC, which displays a somewhat lower ability to form aggregates [69]. Upon addition of quadruplex structures (human telomeric repeat, c-kit1, c-myc, bcl-2), disappearance of the absorption band of the H-aggregate and appearance of the monomer band ($\lambda_{abs} = 580$ nm) were observed, accompanied by a 1,000-fold fluorescence enhancement. At the same time, MTC H-aggregates are almost not dissociated in the presence of ss DNA and ds DNA (Fig. 9).

2.1.2 Quadruplex-Template-Directed Synthesis of Cyanine Dyes

In an elegant approach designed by S. Ladame et al., quadruplex probing was efficiently performed using a biocompatible fluorogenic reaction generating cyanine dyes upon binding of two PNAs, complementary to the quadruplex flanking sequences, and functionalized by nonemissive hemicyanine moieties (Fig. 10) [70]. Upon hybridization of the PNAs with a folded quadruplex structure, the two hemicyanine derivatives are close enough to react, yielding fluorescent trimethine dyes **2** (Fig. 11).

Since the reactants are non-fluorescent when the PNAs are either unbound or bound to unfolded single stranded DNA (no close proximity of reactants), this system offers a high signal-to-noise ratio. Increase in fluorescence intensity in the presence of quadruplex by factors of 45 or 40 were observed for the symmetric cyanine **2a** ($\lambda_{ex} = 542$ nm, $\lambda_{max} = 563$ nm) and the asymmetric cyanine **2b** ($\lambda_{ex} = 562$ nm, $\lambda_{max} = 606$ nm), respectively (Fig. 12).[10] No fluorescence was observed at low PNA concentration and high calf thymus DNA concentration.

[10] Experiments performed in phosphate buffer pH 7.4 with 100 mM K^+.

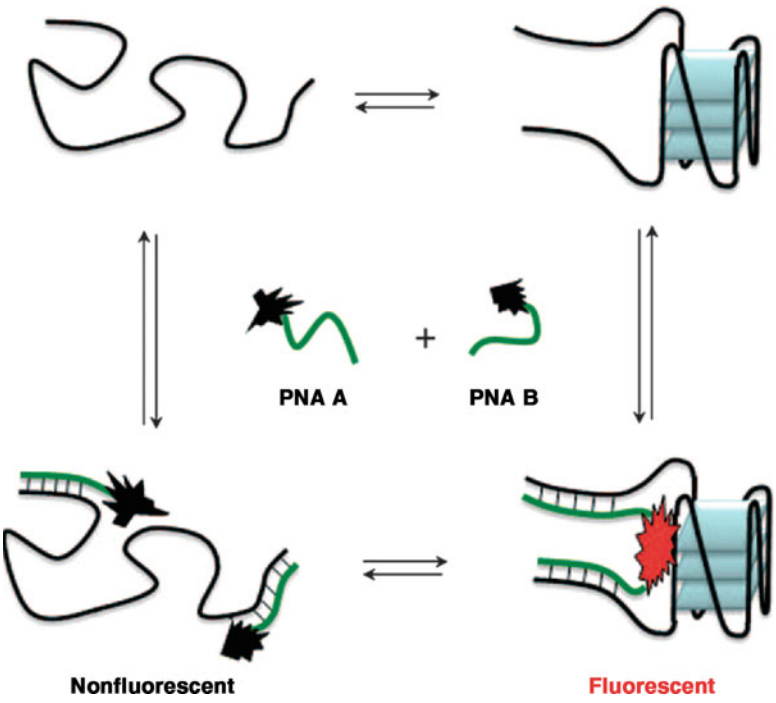


Fig. 10 DNA-templated synthesis of cyanine dyes. Reprinted with permission from [70]. Copyright 2010 John Wiley and Sons

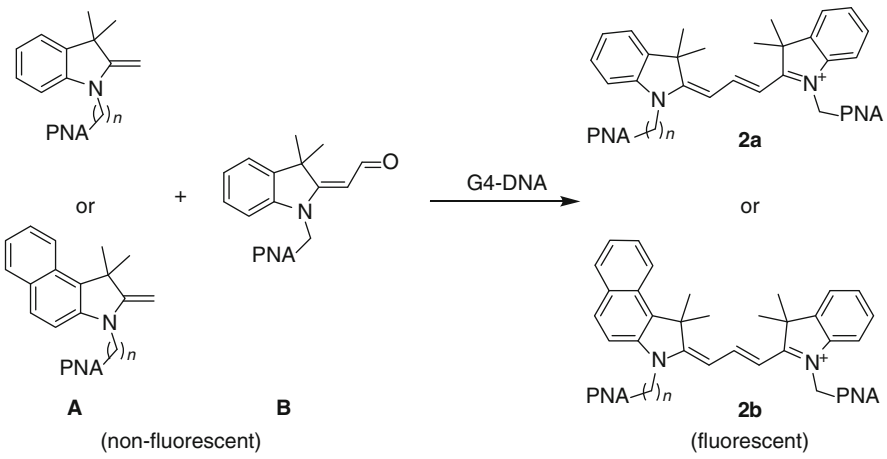


Fig. 11 Trimethine cyanine dyes generated upon PNA binding to G-quadruplex template

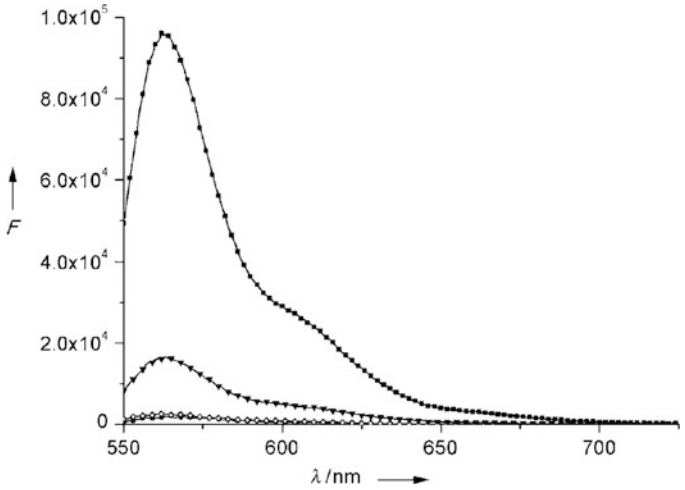


Fig. 12 Fluorescence emission spectra of a mixture of two functionalized PNA (500 nM each) in absence (*circles*) and presence of 500 nM of target quadruplex (*squares*), ss DNA (*diamonds*), and a randomized flanking sequence quadruplex (*triangles*). Reprinted with permission from [70]. Copyright 2010 John Wiley and Sons

Also, the PNA sequence complementarity allows one to target specific quadruplex-forming sequences, and hence potentially to probe single quadruplexes.

This system has been applied for the dual sensing of hairpin and quadruplex DNA structures through the templated generation of pentamethine and trimethine cyanines, respectively [71]. Their resulting fluorescent properties were sufficiently different ($\lambda_{ex} = 540$ and 625 nm, respectively) to detect these structures simultaneously, but with a significant fluorescence background (about 50%).

Although this strategy needs further improvements and in particular will have to overcome the issue of poor cellular uptake of PNAs, it represents a highly interesting alternative to the use of small fluorescent molecules. In particular, this approach based both on sequence and structure recognition holds the key to the specific detection of a unique type of quadruplex DNA.

2.1.3 Benzimidazole Derivatives

Hoechst 33258 (Fig. 13) is a bis-benzimidazole derivative, binding preferentially to minor grooves of AT-rich domains of ds DNA, routinely used for staining nuclear DNA [72]. The fluorescence enhancement observed upon binding, characterized by a large Stokes shift, is explained by the restriction of the dye mobility and its isolation from water [73]. Interaction of Hoechst 33258 with c-myc quadruplex has been examined by Maiti [74] and Patel [75]. As with duplex DNA, the dye is rotationally restricted and isolated from water and may interact in multiple ways (loop, groove, or quartet binding). Hoechst 33258 was used as an alternative to TO in the G4-FID method to assess the compounds absorbing around 500 nm, since its

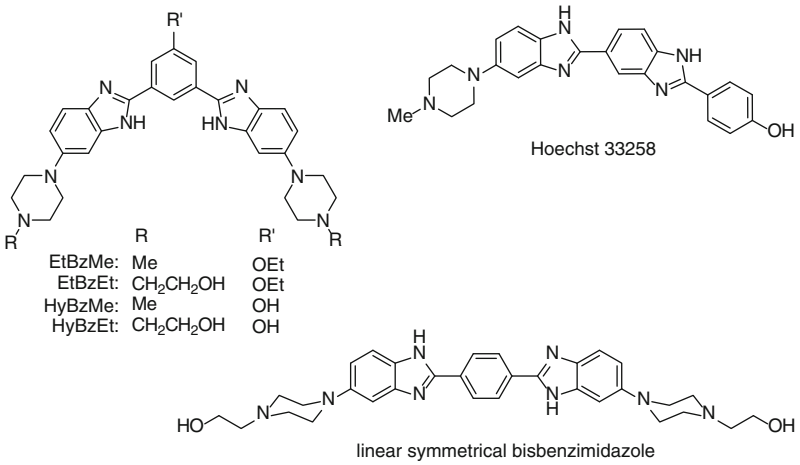


Fig. 13 Bis-benzimidazole quadruplex ligands

absorbance and fluorescence spectra are blue-shifted as compared to TO (λ_{ex} = 343 nm, λ_{max} = 473 nm upon binding to the c-kit2 quadruplex; Fig. 3) [44] (see footnote 1). Of note, the fluorescence enhancements observed with a set of quadruplex structures are dramatically lower (7–19-fold) compared to duplex DNA (36-fold), and the same trend is observed with affinity constants (log K_a ranging from 5.4 to 6.1 for G4 structures vs 6.4 for duplex DNA) [44, 74], thereby confirming that Hoechst definitely binds better to duplex than to quadruplex DNA.

Structurally related phenylene-bisbenzimidazole derivatives (Fig. 13) with a more pronounced crescent shape were synthesized to target quadruplex structures via π-stacking on G-quartets. These compounds display dissociation constants in the range (0.7–5.3) × 10^5 M^{-1} with the *Tetrahymena termophilia* quadruplex d[$(T_2G_4)_4$] [76]. The highest fluorescence enhancement (around 50-fold for a dye-to-DNA ratio of 7.7) was obtained with the EtBzEt derivative, whilst the linear analogue had a poor affinity. These studies demonstrate that duplex minor groove-binding ligands, when suitably modified, may switch their preference from duplex to quadruplex [77].

2.1.4 Ethidium Derivatives

Ethidium bromide (EtBr, Fig. 14) is a classical nucleic acid staining agent, interacting with duplex DNA via intercalation [78]. Fluorescence enhancement upon binding could be explained by protection from the excited-state deprotonation of the solvent [79]. EtBr also binds to triplex DNA and, to a lesser extent, to quadruplex DNA by external stacking [30, 80]. First fluorescence measurements of EtBr in complex with tetramolecular quadruplex DNA T_4G_4 were published over two decades ago, where it was found slightly less fluorescent than in EtBr–duplex

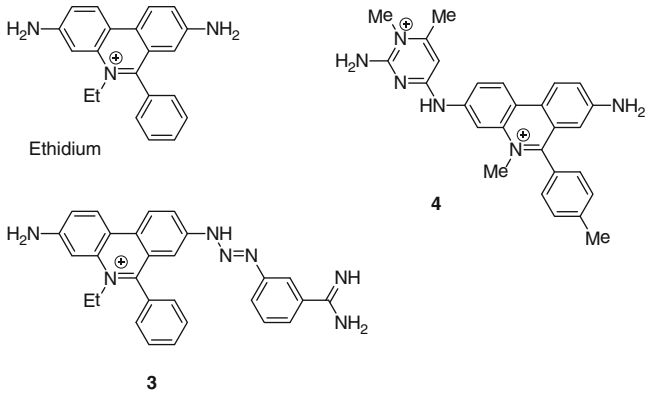


Fig. 14 Structures of ethidium and derivatives thereof

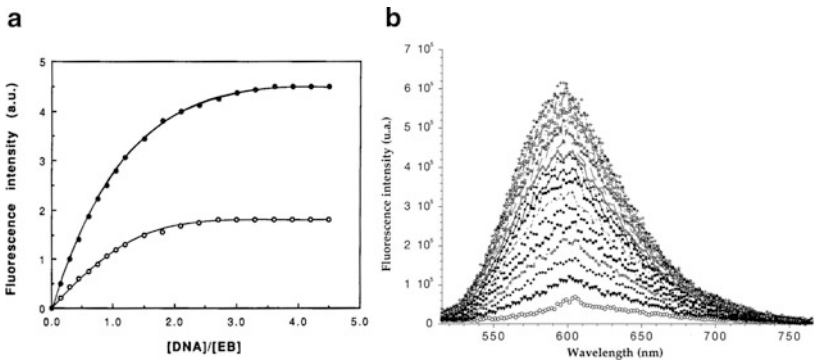


Fig. 15 (**a**) Fluorescence intensity of EtBr (10 μM; $\lambda_{ex} = 510$ nm, $\lambda_{max} = 597$ nm) upon addition of duplex (*filled circles*) and quadruplex (*open circles*) DNA. Reprinted with permission from [81]. Copyright 1992 American Chemical Society. (**b**) Fluorescence titration of compound **4** (400 nM) in the presence of the quadruplex-forming oligonucleotide 28G (40 nM–1.8 μM), $\lambda_{ex} = 500$ nm. *Circles* indicate emission spectrum without 28G. Adapted with permission from [82]. Copyright 2001 Oxford University Press

DNA complexes (Fig. 15a; $\lambda_{ex} = 510$ nm, $\lambda_{max} = 590$ nm) [81]. Further studies confirmed the weaker fluorescence enhancement and hypochromic shift of EtBr in the presence of quadruplex DNA as compared with duplex or triplex DNA [83].

Ethidium derivatives were prepared with the aim of improving their quadruplex-binding capabilities while retaining interesting fluorescence properties [82, 84]. Two of them (**3** and **4**, Fig. 14) display strong fluorescence increase upon binding to the human telomeric quadruplex structure (cf. Fig. 15b). The increase of fluorescence quantum yield (from 0.06 and 0.04 to 0.25 and 0.20, respectively) is complemented by the shifts of absorbance and emission wavelengths ($\lambda_{ex} = 477$ and 464 nm to 508 and 494 nm and $\lambda_{max} = 597$ and 603 nm to 595 and 592 nm,

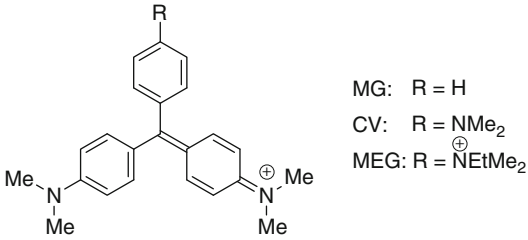


Fig. 16 Triphenylmethane dyes

respectively).[11] It was also observed that binding of **4** to a parallel-stranded quadruplex structure d[$(T_2G_4T)_4$] leads to a lower fluorescence increase than with an antiparallel structure, probably because the dye is more exposed to the solvent. The two derivatives were successfully used to stain quadruplex DNA in agarose gels, but with a 100-fold lower sensitivity than the cyanine dye DODC.

2.2 Triphenylmethane Dyes

Malachite green (MG), crystal violet (CV), and methyl green (MEG) (Fig. 16) belong to the class of triphenylmethane (TPM) dyes that have been known for more than a century. The characteristic feature of TPM dyes is the existence of several conformers in the fundamental state and the formation of a nonemissive twisted intramolecular charge transfer (TICT) state resulting from diffusional mutual rotation of the phenyl rings [85, 86]. Consequently, the fluorescence quantum yield of the free dyes in organic solvents is usually very low ($\approx 10^{-5}$) but increases significantly with the viscosity of the solvent due to restriction of intramolecular rotations [87]. Likewise, the fluorescence of TPM dyes (malachite green and derivatives) has been shown to be enhanced upon binding to RNA aptamer matrices in which the molecule is immobilized [87]. Two TPM dyes have been studied recently for their interaction with quadruplex-forming sequences. MG was shown to bind the G-rich sequence d[$(G_2T)_{13}G$] which resulted in a 100-fold increase in dye fluorescence (Fig. 17) [88]. Again, this phenomenon has been attributed to increased molecular rigidity of the dye when bound to the quadruplex DNA structure. Interestingly, the fluorescence enhancement is much lower (14-fold) with single-stranded and double-stranded oligonucleotides used as controls. However, the interaction appears rather complex and the dye interacts only at low concentration in oligonucleotide, thus suggesting the recognition of an intramolecular quadruplex. The structural information is scarce as the quadruplex that can be formed by this unusual sequence is only characterized by thermal difference spectra (TDS) [89]. Nevertheless, a 70-fold increase in fluorescence is also observed with the more classical $(G_4T_4)_4$ quadruplex but at high concentration. Follow-up studies

[11] Experiments performed in 10 mM cacodylate buffer pH 7.3 with 100 mM K^+.

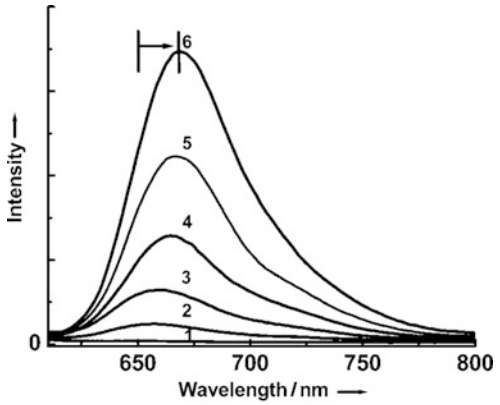


Fig. 17 Enhancement of the fluorescence intensity of MG (1 μM) in an aqueous solution containing phosphate buffer (5 mM) at pH 7 with d[$(G_2T)_{13}G$] at increasing concentrations: (1) 0 μM, (2) 6 μM, (3) 60 μM, (4) 96 μM, (5) 108 μM, (6) 120 μM. Reprinted with permission from [88]. Copyright 2007 John Wiley and Sons

by Shen et al. using several TPM dyes (CV, MG, MEG, Fig. 16) and a panel of well-defined quadruplex-forming sequences (intramolecular and tetramolecular G-quadruplexes from the telomeric human sequence [hum21, hum12] or from *Oxytricha nova* [Oxy28]) have shown that energy transfer spectra of TPM dyes can serve to discriminate intramolecular and tetramolecular quadruplexes from other DNA forms [90, 91]. Energy transfer is a reliable indicator of π-stacking interactions and it is likely that the TPM dyes interact with external G-quartets. CV and MEG bind better than MG with dissociation constants in the low micromolar range ($K_d = 0.036$ μM for CV; 0.026 μM for MEG) [92].[12] However, a competition dialysis experiment indicated a rather moderate binding selectivity for intramolecular quadruplexes with respect to double-stranded DNA (ct DNA) and no preference for the tetramolecular telomeric quadruplex (hum12). Nonetheless, CV was subsequently proposed for the development of fluorescent switch-on probes for homogeneous detection of potassium ions [93], for i-motif [94], and the development of a G-quadruplex-based turn-on fluorescence assay for the $3' \rightarrow 5'$ exonuclease activity [95].

It is worth noting that, in spite of significant fluorescence enhancements (around 70–100-fold), the fluorescence quantum yield of TPM dyes globally remains low (around 10^{-3}), which is a limiting factor for fluorescence applications using routine equipment (for comparison, fluorescence enhancements of G4-bound TO are in the 1,000–3,000-fold range). This, together with the poor ability of the series to discriminate between different biomacromolecules [87, 96], minimizes the practical interest of TPM dyes for quadruplex sensing applications.

[12] [CV] fixed at 15 μM in 25 mM Tris–HCl (pH = 7.0), the DNA concentrations (strand concentration) at intersection points on the titration curves is 8 ± 1 μM for Hum21.

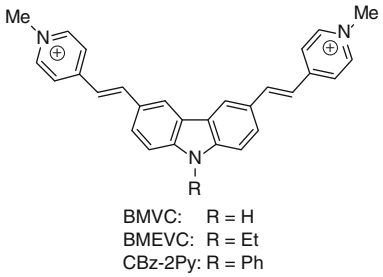


Fig. 18 Structure of carbazole derivatives

2.3 Carbazoles

In 2003, the group of T.-C. Chang reported the synthesis of a novel carbazole derivative featuring two vinylpyridinium branches in positions 3 and 6 (Fig. 18). This compound (called BMVC) was shown to stabilize significantly the human telomeric quadruplex ($\Delta T_{1/2} = +13$ °C measured by CD-melting).

Interestingly, the binding of the dye was accompanied by a strong fluorescence enhancement evaluated to be around two orders of magnitude as compared to the free dye that is poorly fluorescent [97, 98]. BMVC belongs to the family of cationic styryl dyes that are characterized by fast deactivation of their excited state due to rotational vibrations around the vinylic bond. This well-known phenomenon results in a strong quenching of the fluorescence of the unbound compounds whilst their fluorescence is restored upon immobilization in viscous solvents (typically glycerol) or in polymeric biological matrices such as DNA. Subsequently the photophysical properties of this dye were thoroughly investigated in correlation with its properties for quadruplex DNA recognition, focusing mainly on the human telomeric quadruplex. Although BMVC binds the human telomeric quadruplex d[$(T_2AG_3)_4$] with a nanomolar K_d [99], it also seems to bind rather well to duplex DNA structures like calf thymus DNA and linear short duplexes, although the affinity for the latter forms has not been determined. Nonetheless, in certain conditions, the dye appears to have a clear preference for the human telomeric quadruplex-forming sequence d[$(T_2AG_3)_4$] as shown by a competition assay carried out on gel electrophoresis (Fig. 19) [100]. In this assay the quadruplex is selectively stained when a mixture of duplex and quadruplex is mixed with 0.1 μM dye (called prestaining). Subsequent post-staining, i.e., incubation of the gel after migration in a solution of BMVC at 10 μM, enables labeling of both duplex and quadruplex DNA forms. The dye reveals an extreme sensitivity since as little as 0.2 pmol of quadruplex can be detected in the gel.

Another interesting feature of the BMVC fluorescence is the difference in the maximum emission wavelength when bound to quadruplex or duplex DNA. Indeed, a strong blue shift is observed in the presence of duplex DNA ($\lambda_{max} = 545$ nm) as compared to the red-shifted emission in the presence of quadruplex ($\lambda_{max} = 575$ nm).

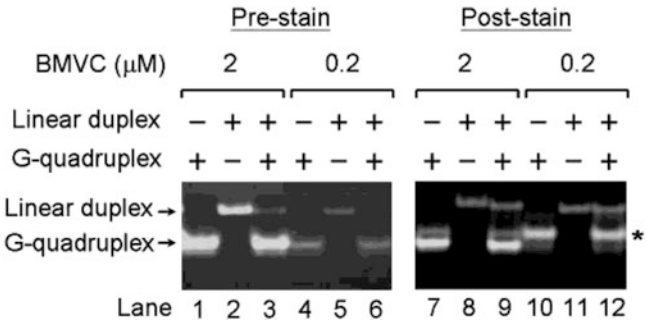


Fig. 19 Binding preference of BMVC to quadruplex DNA. 20 μM of quadruplex DNA $d(T_2AG_3)_4$ or linear duplex $(d[ATGCGCA_2T_2GCGCAT])_2$ were incubated with BMVC (2 or 0.2 μM) and analyzed on a 20% polyacrylamide gel (prestain, lanes 1–6). The same gels were then poststained with 10 μM BMVC to visualize the position and level of DNA loaded (poststain, lanes 7–12). An *asterisk* indicates position of the free quadruplex DNA. Reprinted with permission from [100]. Copyright 2004 American Chemical Society

This is clearly seen by the different colors (yellow to orange) of the fluorescence emission of the DNA–dye complexes both on gel and in solution (Fig. 20a). Although this is not commented on by the authors, this difference indicates that BMVC is in a less polar environment when bound to duplex than when bound to quadruplex which reflects better shielding from water upon interaction with the double helix structure.

This ability of BMVC to signal differently duplex and quadruplex structures was tentatively applied for detecting the presence of quadruplex DNA at the level of chromosomes. Chromosomes in metaphases were observed by fluorescence microscopy and the fluorescence was collected from various regions. Telomeric proximal regions showed fluorescence emissions around 565 nm whereas the other chromosomal regions showed fluorescence emission around 545 nm (Fig. 20b). On this basis, the authors concluded on the presence of quadruplex structures at telomeres. However, both the specific conditions required for this experiment (the red-shifted signal is seen only at short staining time, i.e., 5 min and very low dye concentration of 2 nM) and the associated technical difficulties, i.e., the precision in term of spatial resolution and imaging contrast, are not addressed.

In a follow-up study, the lifetime of the excited state of BMVC in the presence of various DNA forms was measured. A longer lifetime was observed in the telomeric quadruplex (1.96 ns with $d(T_2G_3)_4$ in Na^+ conditions), as compared to the linear duplex (1.46 ns with $(d[ATGCGCA_2T_2GCGCAT])_2$). However, in K^+ conditions the lifetime of the dye bound to the quadruplex was also shorter (1.56 ns with $d[(T_2AG_3)_4]$ in K^+ conditions). The authors conclude that the long lifetime is characteristic of the antiparallel form whilst the short component is due to the parallel form, assuming that the latter dominates in potassium conditions. Two-photon-excited fluorescence lifetime imaging (2PE-FLIM) was then performed with the aim of detecting the presence of quadruplexes on chromosomes in

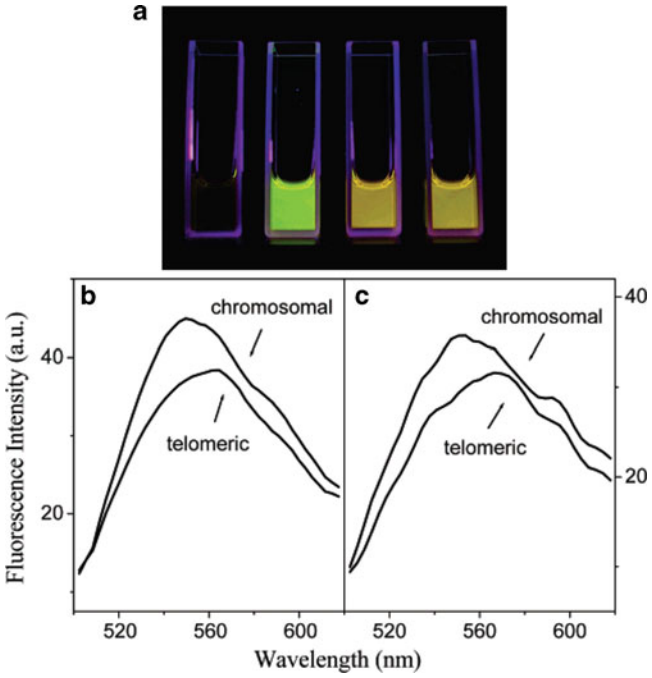


Fig. 20 (**a**) Fluorescence of solutions of BMVC free and in the presence of duplex or quadruplex DNA; from left to right: BMVC alone, in the presence of a linear duplex (d[ATGCGCA$_2$T$_2$GCGCAT])$_2$, and d(T$_2$AG$_3$)$_4$ quadruplex in Na^+ and K^+ conditions, respectively. Adapted with permission from [101]. Copyright 2006 American Chemical Society. (**b**, **c**) Fluorescence collected on chromosomes in metaphases of human CL1-1 (**b**) and Ca9-22 cells (**c**). Adapted with permission from [100]. Copyright 2004 American Chemical Society

metaphase. The long-lifetime component was observed in the telomeric regions arguing again for the presence of quadruplex in its antiparallel form although potassium is predominant in the cells (Fig. 21). Consistent with the previous observation on chromosomes using the red-shifted signature [100], the quadruplex-specific long-living component is time and concentration dependant, suggesting that BMVC associates first to the (presumed) very low amount of quadruplexes present in the genome and then redistributes on the massive excess of duplex domains upon equilibration. Kinetic studies would thus be required to fully understand the behavior of BMVC and to conclude firmly on the presence of quadruplex structures at telomeres.

In parallel studies, BMVC was used to explore the interconversion between the antiparallel and the parallel forms of the telomeric quadruplex both on short (24 bases) and long (78 bases) sequences in combination with CD spectroscopy. The dye exhibits a strong ICD signal upon interaction with both parallel and antiparallel quadruplexes [99]. Finally, molecular dynamics studies have shown that BMVC likely binds to the basket Na^+-form of the telomeric quadruplex by external

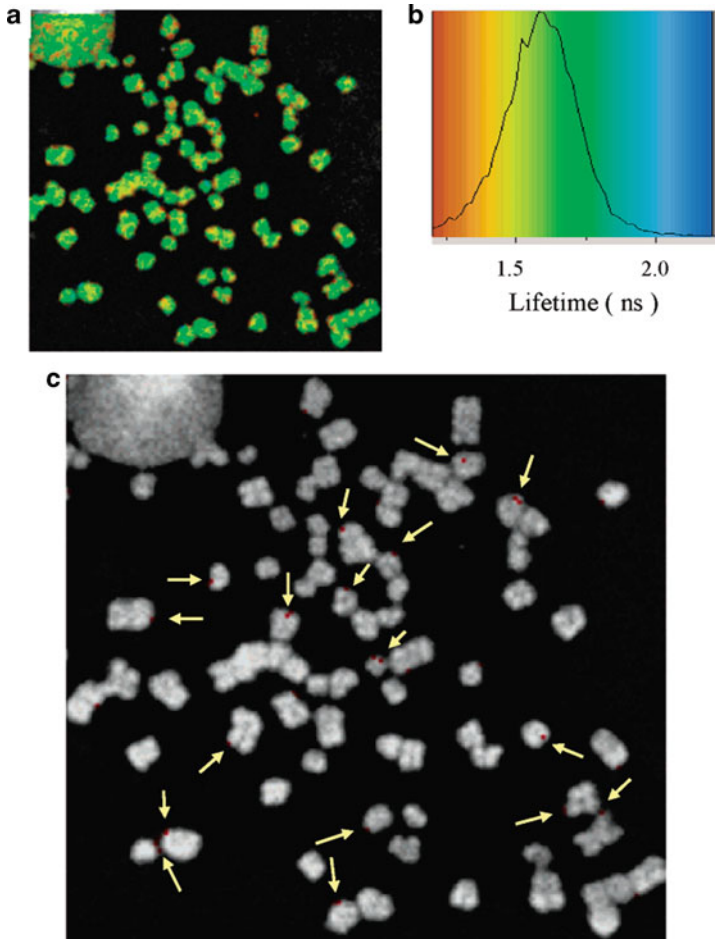
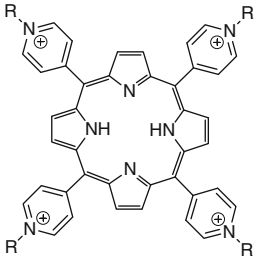


Fig. 21 2PE-FLIM micrograph of BMVC-stained metaphase chromosomes of nasopharyngeal carcinoma KJ-1 cells. (**a**) A typical 2PE-FLIM image of metaphase chromosomes of nasopharyngeal carcinoma KJ-1 cells stained with 0.1 μM BMVC for 90 s. (**b**) The fluorescence lifetime distribution of the BMVC-stained chromosomes. (**c**) The color-coded image of the metaphase chromosomes is represented by discrete timemode at 1.85 ns. *Red* designates mode 1 resulting from the interaction with antiparallel G-quadruplex structures, while *white* designates mode2 due to the interaction with other DNA structures. The *arrows* indicate mode 1 for visualization. Adapted with permission from [101]. Copyright 2006 American Chemical Society

π-stacking on the two G-quartets combined with interactions with the diagonal and parallel surrounding loops [99]. The strong involvement of loops in the interaction would explain the binding preference for the antiparallel telomeric quadruplex suggested by the initial gel analysis.

More recently, reports by other groups concerning *N*-alkyl [102] or *N*-aryl analogues [103] of BMVC (Fig. 18) have been published. The first study based



TMPyP4: R = Me
TrPyP4: R = *n*-Pr

Fig. 22 Structures of TMPyP4 and TrPyP4

on modeling suggests that the bis(vinylpyridinium)carbazole core could fit nicely in the grooves of the antiparallel form of the telomeric quadruplex. The second study demonstrates that the *N*-phenyl derivative of BMVC (Cbz-2py) binds both duplexes and various intramolecular quadruplexes (telomeric, c-myc, c-kit). However, this compound exhibits a strong and preferential fluorescence enhancement in AT-rich duplexes likely resulting from a net preference for duplex minor groove binding. This would tend to indicate that BMVC is better in terms of fluorescence selectivity for quadruplex over duplex structures.

The major interest of the studies carried out by T.-C. Chang et al. is to have made use of the particular photophysical characteristics of BMVC to propose diverse fluorescent-based approaches that may allow signaling of quadruplex-containing regions at the level of chromosomes, although the probe has a relatively poor structural preference for quadruplexes. Thus this work sets the stage for future possible design of experiments based on probes with optical features (fluorescence lifetime, emission wavelength) specific for quadruplexes and combined with current advanced technologies (one- and two-photon FLIM mapping).

2.4 Porphyrins

Porphyrins have been thoroughly investigated for their interaction with quadruplex DNA following the pioneering work of L. Hurley et al. on the 5,10,15,20-tetrakis(1-methyl-4-pyridyl)porphine called TMPyP4 (Fig. 22) [104]. Porphyrins are fluorescent compounds but exhibit highly variable quantum yields as a function of the substituents born by the tetrapyrollic scaffold and whether they are in the free-base form or metallated. For instance, most of the pyridinium series developed to interact with quadruplexes [105–107] show either poor fluorescence variations or fluorescence attenuation due to electron transfer from guanines to the porphyrin excited state, likely favored by the strong electron-acceptor character of the pyridinium units (see below Sect. 3, Light-off probes). However, fluorescence

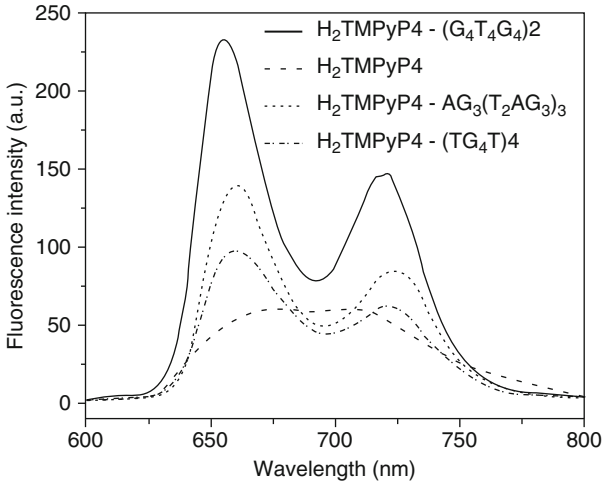


Fig. 23 Fluorescence emission spectra of TMPyP4 (4 μM) and its complexes with $(TG_4T)_4$, $AG_3(T_2AG_3)_3$, and $(G_4T_4G_4)_2G$-quadruplexes (20 μM). Adapted with permission from [105]. Copyright 2006 American Chemical Society

increase of TMPyP4 has been mentioned in several cases. Fluorescence quantum yield of free TMPyP4 in water is moderate ($\Phi_F = 0.05$) with a typical broad dual-band pattern exhibiting two maxima in the NIR region (660 and 720 nm). Upon association with tetramolecular and bimolecular quadruplexes $(d[TG_4T])_4$, $(d[G_4T_4G_4])_2$ and with the intramolecular telomeric sequence $d[AG_3(T_2AG_3)_3]$ the emission intensities increase about 1.5–2.5 times [107] in Na^+ conditions, displaying a well-structured spectrum (Fig. 23).[13] In contrast, a strong decrease (>70%) is reported with intramolecular and dimeric hairpin quadruplexes from telomeric sequences $d[TAG_3(T_2AG_3)_3T]$ and $d[T_2AG_3T_2AG_3)_2]$ in K^+ conditions (10 mM K^+) [106]. Finally, a slightly different pattern was observed upon interaction with the parallel-quadruplex-forming $d[G_4(T_2AG_3)_4]$, showing an increase of the band at 658 nm and a decrease of the band at 718 nm, which might be explained by modification of the transition probability of the two vibrational states [108]. On the whole, all these studies suggest that the fluorescence of TMPyP4 is highly dependent on the structure of the bound quadruplex.

In the follow-up studies the *N*-propyl analogue TrPyP4 showed only slight changes of fluorescence intensity upon addition of various G4 sequences; however, as in the case of TMPyP4, binding to G4 DNA induced a better resolution of the emission spectrum with the appearance of clear structured vibrational contributions [109–111].

A series of porphyrin-anthraquinone dyads **5** was prepared [112] with various chain length (from 1 to 10 methylene units; Fig. 24). Compounds with the longer

[13] All solutions contain 10 mM Tris–HCl, 1 mM Na_2EDTA, and 100 mM NaCl at pH 7.5.

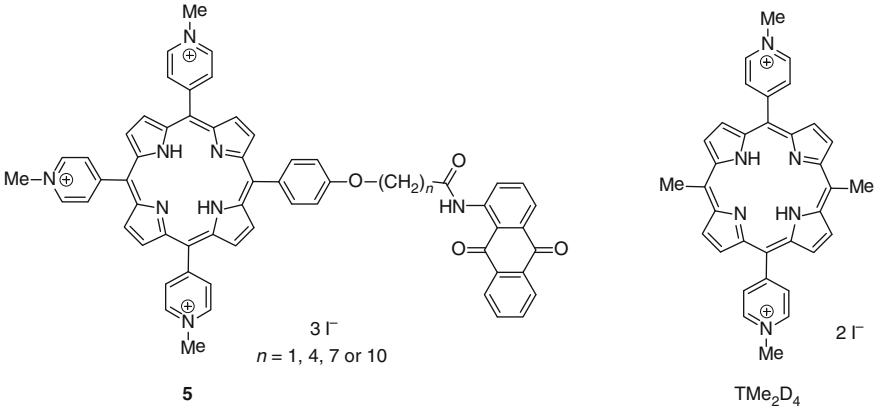


Fig. 24 Structures of anthraquinone conjugate **5** and TMe_2D_4

chain length (7 or 10 methylene units) exhibit a higher affinity for $(d[TG_4T])_4$ ($\Delta T_{1/2} = 12.8$ and 11.2 °C, respectively, as determined by UV-melting) than those with the shorter linkers ($\Delta T_{1/2} = 3.0$ and 6.4 °C for derivatives with 1 and 4 methylene units, respectively). Upon addition of increasing concentration of $(d[TG_4T])_4$ the derivatives with 4, 7, and 10 methylene units in the chain display similar fluorescence emission spectra, peaking at 665 nm with an enhanced intensity (~30%), whilst such increase is not observed with the shortest-chain analog.

The metallated forms of the cationic bis(pyridinium)porphyrin TMe_2D_4 (Fig. 24) exhibit more interesting fluorescence properties. In particular, $Cu(TMe_2D_4)$ was reported to exhibit a strong fluorescence enhancement (14-fold) with an infrared-shifted emission peaking at 850 nm in the presence of tetramolecular quadruplexes $(d[T_nG_4T_n])_4$. Interestingly, the emission increases in intensity as the number of thymines increases from 1 to 3, likely due to enhanced shielding from water that is a strong quencher of copper porphyrins. Worth mentioning is the fact that the emission of Cu(TMPyP4) is lower and furthermore is not affected in these conditions.

Anionic mesoporphyrins bearing carboxylate groups (Fig. 25) have also been investigated as fluorescent probes for G4-DNA. The anionic *N*-methyl-mesoporhyrin IX (NMM) exhibits a marked selectivity for quadruplex structures as compared to duplex with a good affinity, only recently quantified ($\Delta T_{1/2} = 14$ °C determined by FRET-melting) [113]. In early studies the fluorescence of NMM was shown to be strongly enhanced in the presence of dimeric and monomeric G4-forming sequences $(d[G_4T_4G_4])_2$ and TBA, whereas no variation was observed in the presence of calf-thymus DNA [114].

Protoporphyrin IX (PPIX) is known as a photodynamic agent in cancer therapy, able to interact with various proteins. On the tracks of NMM, binding of PPIX against various G4-DNA (TBA, human telomeric, or Oxy28) has been investigated [115]. The emission spectrum of PPIX is typical of a porphyrin compound with a dual band in the 600–700 nm range. The fluorescence enhancement efficiencies observed with

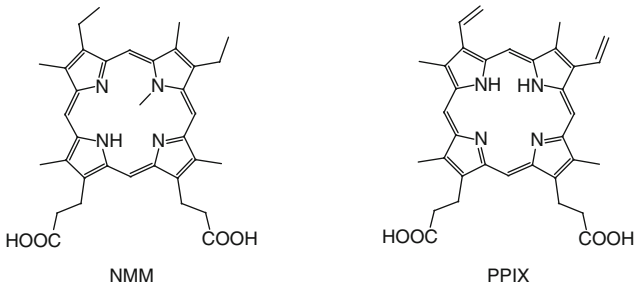


Fig. 25 Structures of *N*-methyl-mesoporhyrin IX (NMM) and protoporphyrin IX (PPIX)

different G4-DNA are from 16.2 to 17.9-fold, whilst the binding constants are rather different (K_d from 0.04–50 μM). In the same work, PPIX has also been used as a "sensor" of parallel quadruplexes formed in K^+ conditions as the fluorescence increases rapidly when the K^+ concentration rises from 2 to 100 mM. The fluorescence properties of Zn-PPIX have also been used to monitor the lead-driven interconversion between a duplex and a G4 form acting as components of a DNA nanodevice [116].

2.5 *Probes Operating via Aggregation-Induced Emission*

Aggregation-induced emission (AIE) is a process observed in an unusual fluorogen system, in which aggregation enhances the fluorescence [117]. It has been found that these systems are nonemissive in dilute solutions, but become highly luminescent when the molecules are aggregated in concentrated solutions. 1,1,2,2-Tetrakis [4-(2-triethylammonioethoxy)phenyl]ethene (TTAPE, Fig. 26) can be considered an archetypal AIE luminogen. Structural analysis reveals that unlike conventional luminophores, such as the disk-like planar aromatic molecules (for example perylene derivatives; cf. Sect. 3.3), TTAPE is a propeller-shaped nonplanar molecule. In a dilute solution, four phenyl rotors in TTAPE undergo dynamic intramolecular rotation about its ethene stator, which nonradiatively annihilates the excited state and renders the molecule nonfluorescent. In the aggregates, the molecules cannot efficiently pack through a π–π stacking process due to their propeller shape; at the same time, the intramolecular rotations of the aryl rotors are greatly restricted owing to the physical constraint. This restriction of intramolecular rotation (RIR) blocks the nonradiative pathway and opens up the radiative channel. As a result, the fluorogen becomes emissive in the aggregate state.

AIE fluorogens are useful analytical tools and have been judiciously utilized as sensitive and selective chemosensors and bioprobes of the "light-up" type. In many of these sensing systems, the aggregation of the free AIE molecules in solution is triggered by electrostatic attraction and coordinative complexation to the analyte [118, 119]. Thus, the emission of the TTAPE that is weak in aqueous buffer

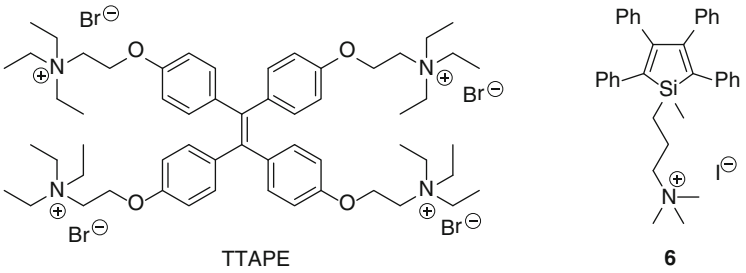


Fig. 26 Probes operating via aggregation-induced emission

($\Phi_F \approx 0.5\%$) is restored in the presence of certain quadruplex-forming sequences [118]. In particular, the largest fluorescence enhancement (45-fold) was obtained for the human telomeric DNA sequence d[$G_3(T_2AG_3)_3$] in the presence of K^+ ions, whereas almost no variation was induced by the same sequence in the presence of Na^+, indicating selective recognition of the K^+ conformers. Subsequently, more modest but still significant fluorescence increase (10–12-fold) was observed with the c-myc derived sequence d[$TG(AG_3TG_3T)_2A_2$] and with the telomeric sequence from *Oxytricha nova* d[$G_4(T_4G_4)_3$] (Fig. 27). This also enables selective staining of the telomeric quadruplex in gel [119]. Of note, TTAPE also binds to duplex DNA but the emission of TTAPE-G4 and TTAPE-ds DNA complexes display significant differences: the emission in quadruplex is twice as intense and is characterized by a bathochromic shift by 20 nm. This observation is presented as a way to discriminate between the two DNA forms [118]. A molecular modeling experiment indicates that the dye may be stacked on the surface of the quadruplex structure in a loop–groove interaction based on multiple electrostatic interaction with phosphates, which impedes the intramolecular rotations and favors the radiative relaxation [119]. The binding constant determined by fitting the fluorimetric titration curve obtained with the human telomeric G4-DNA in K^+ is rather modest ($K_a = 7.8 \times 10^5\ M^{-1}$).[14] A very close value was found by calorimetric titrations (ITC) ($K_a = 2.4 \times 10^5\ M^{-1}$), which also enabled to determine a remarkably low affinity constant for the same sequence in Na^+ conditions ($K_a = 4.3 \times 10^2\ M^{-1}$) [119]. Additional information about the changes in the fluorescence emission of TTAPE in the presence and in the absence of DNA was obtained from the time-resolved fluorescence spectra. This technique offers valuable information about the interactions of the fluorogens in the excited state with the environment, by measuring the fluorescence lifetime. In buffer solution and in absence of DNA, TTAPE decays in a single-exponential way with a very short lifetime of 20 ps, which is consistent with the involvement of a nonradiative process of de-excitation. The fluorescence lifetime is lengthened to a few nanoseconds (3–4 ns) when the human telomeric

[14] [TTAPE] = 4.5 μM, [K^+] = 0.15 M in 5 mM Tris–HCl buffer; $\lambda_{ex} = 350$ nm.

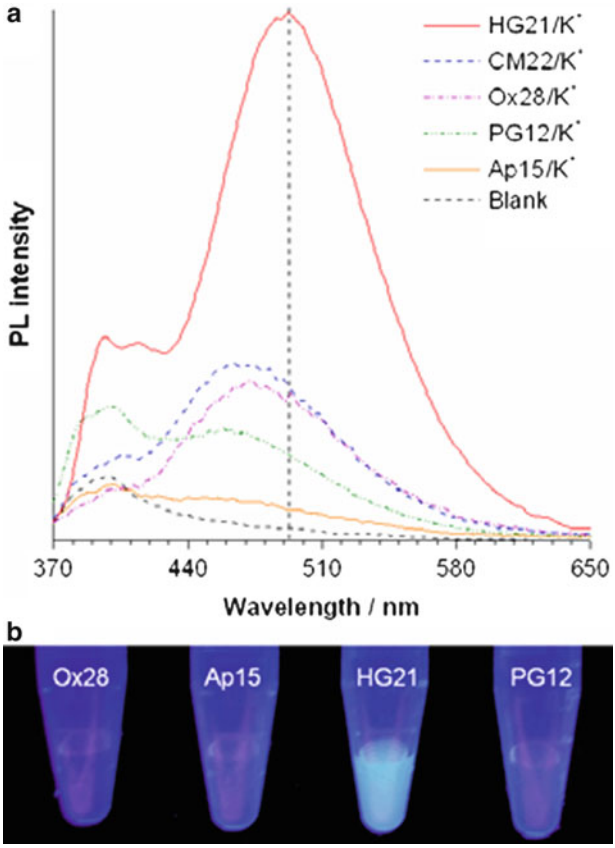


Fig. 27 (**a**) Photoluminescence spectra of buffer solutions of TTAPE in the presence of K^+ and G-quadruplexes with different conformations. [TTAPE] = 4.5 μM, [DNA] = 9 μM, [K^+] = 0.5 M; λ_{ex} = 350 nm. (**b**) Photographs of TTAPE solutions in the presence of different G-quadruplexes taken under UV illumination (365 nm). Reprinted with permission from [119]. Copyright 2010 John Wiley and Sons

quadruplex is added to the solution, and the fluorescence decay curve is better fitted by a double-exponential function [119].

TTAPE was also evaluated in cells: the compound penetrates into living HeLa cells and appears nontoxic. However, the dye does not reach the nucleus as the fluorescence is exclusively observed in the cytoplasmic compartment [119].

Another remarkable AIE molecule is the silole **6**, reported for the first time by Tang and co-workers [120]. The fluorescence of silole **6** increases largely upon mixing with DNA, and large fluorescence enhancements were detected for long DNA, indicating that **6** can be used for fluorescence turn-on detection of DNA. More interesting, this behavior enabled following the DNA cleavage process by nucleases [121]. This behavior prompted the authors to exploit silole **6** as a fluorescent light-up probe ($\lambda_{max} \approx 470$ nm) to detect G-quadruplex formation. Compound **6** was employed to study G-quadruplex formation using an exonuclease

I hydrolysis assay. The fluorescence of **6**, in the presence of a duplex containing the telomeric sequence (21-mer) and in absence of K^+, gradually becomes weak by prolonging the enzymatic hydrolysis reaction time. This is due to fragmentation of DNA by exonuclease I, and as a result the aggregation of silole **6** cannot take place efficiently. However, upon addition of KCl, the fluorescence of silole **6** is only slightly affected, which is attributed to the K^+-induced formation of quadruplex that inhibits the hydrolysis reaction. Very similar results have been obtained in the presence of various G-quadruplex forming sequences such as c-myc, c-kit, VEGF, and an intermolecular G-quadruplex $d(G_4T_4G_4)$ [122].[15]

2.6 Miscellaneous

2.6.1 Natural Products

Several aromatic natural products are weakly fluorescent when free in aqueous buffer whilst their binding to quadruplex DNA may significantly increase their fluorescence intensity. Among them are the isoquinoline alkaloids, berberine and palmatine (Fig. 28), which were initially identified as inhibitors of topoisomerases I and II, the latter being the targets of several anticancer drugs used in the clinic [123].

Both compounds include an aromatic moiety with a quaternary bridgehead nitrogen atom, and appear suitable for stacking interactions with the G-quartet. Berberine has a very low quantum yield in aqueous solution ($\Phi_F = 4.5 \times 10^{-4}$) [124], which is increased by a factor of 50 upon addition of quadruplex DNA while the maximum in the fluorescence emission spectrum ($\lambda_{max} = 522$ nm) is slightly blue-shifted by 5 nm. The analysis of the binding isotherms revealed a 1:1 binding model with a binding constant of 1.2×10^6 M^{-1} for berberine binding to human telomeric quadruplex $d[AG_3(T_2AG_3)_3]$ (Fig. 29) [125]. Fluorescence polarization anisotropy measurements of berberine and palmatine bound to G-quadruplex structures revealed an increase of fluorescence anisotropy as compared to the free compounds (0.140 and 0.141 in the complex with G4-DNA vs 0.038 and 0.032 in the free state, respectively), indicative of binding. However, the observed values are lower than those found for the planar analogues (e.g., coralyne, cf. Sect. 3.2), indicating that buckling of berberine and palmatine hinders their interaction with quadruplex DNA [126].

By taking advantage of the berberine fluorescence enhancement upon binding to G-quadruplex, a new label-free and homogenous fluorescent assay for K^+ has been developed. Upon addition of K^+ ions to a mixture containing a G-rich single-stranded DNA (TBA, $d[G_2T_2G_2TGTG_2T_2G_2]$) and berberine, a remarkable enhancement of fluorescence is induced. Interestingly, the method shows a high

[15] 20 μM **6** in 20 mM pH 7.4 Tris buffer, containing 2.5 μM quadruplex in absence or in the presence of 100 mM KCl.

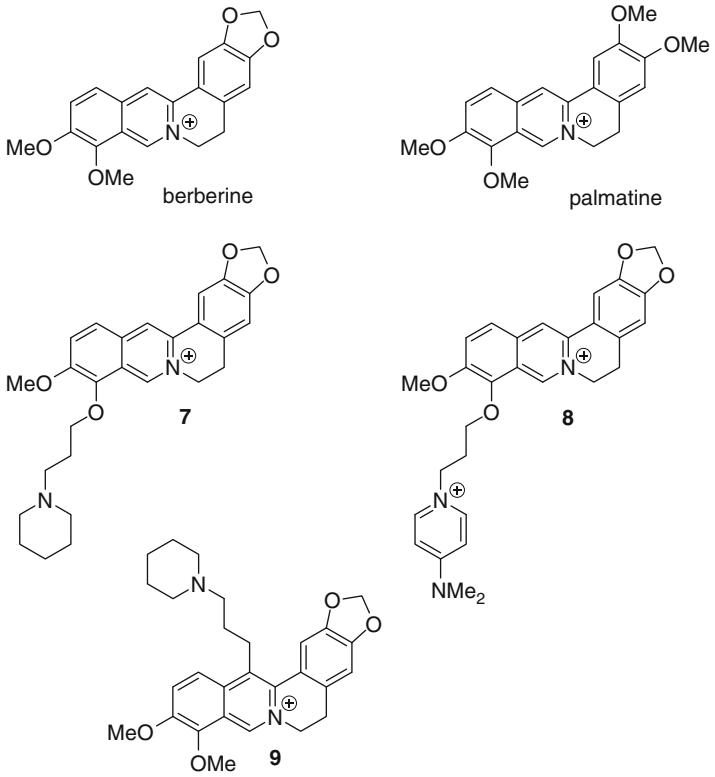


Fig. 28 Berberine derivatives

selectivity for K^+ as detection occurs in the presence of 10,000-fold excess of Na^+ [127]. The introduction of an alkylamino group, such as a 3-(1-piperidino)propyl substituent (compound **7**) or a positively charged aza-aromatic terminal group (compound **8**), into position 9 of berberine improves the binding selectivity for G-quadruplex, while keeping the significant fluorescence "light-up" properties ($\Delta T_{1/2}$ = 28.2 °C for **7** with the telomeric sequence d[$G_3(T_2AG_3)_3$] vs $\Delta T_{1/2}$ = 3.9 °C for berberine, calculated from the CD melting curves at 295 nm) [128].[16] Very similar results were obtained for compound **8** that showed a $\Delta T_{1/2}$ value of 25 °C (obtained from FRET-melting experiments in the presence of F21T) [129].[17,18] Similar behavior was observed for the 13-substituted berberine derivative (**9**) [130].

[16] Measured by CD spectra in 10 mM Tris–HCl buffer (pH 7.4) without metal ions.

[17] $\Delta T_{1/2}$ values of 0.2 μM F21T (21-mer human telomeric sequence Fluo-GGG$(TTAGGG)_3$]-TAMRA) incubated in the presence of KCl 60 mM, compound **2** 2.0 μM.

[18] 10 mM Tris–HCl buffer (pH 7.4) and 100 mM KCl.

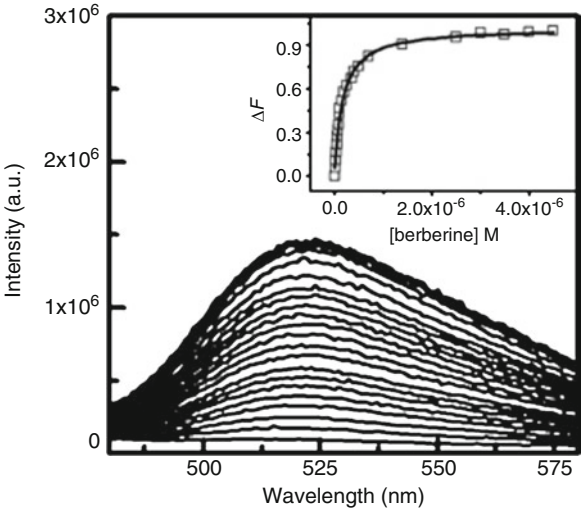


Fig. 29 Fluorescence emission spectra of 0.5 μM berberine in 50 mM MES buffer (pH 7.4) and 100 mM KCl in the absence and presence of successive addition of quadruplex at 25 °C. The *inset* is the plot of ΔF vs quadruplex concentration. Reprinted with permission from [125]. Copyright 2008 John Wiley and Sons

Isaindigotone (Fig. 30) has a structure including a pyrrolo[2,1-*b*]quinazoline moiety conjugated with a benzylidene group. It is a naturally occurring alkaloid isolated from the root of *Isatis indigotica Fort*, a herb commonly used in traditional Chinese medicine. It may be assumed that, like in the case of cyanine dyes (cf. Sect. 2.1), the rotation of the aryl substituent in the isaindigotone scaffold can be hindered upon binding to the quadruplex, leading to enhancement of fluorescence. While isaindigotone itself does not show significant affinity towards quadruplex DNA, derivatives endowed with two aminoalkyl chains (**10**, **11**) were identified as promising G-quadruplex ligands [131]. These two compounds were shown to be potent and selective stabilizers of the telomeric quadruplex ($\Delta T_{1/2} = 21.9$ °C for **10** and 17.6 °C for **11** bound to d[$(G_3(T_2AG_3)_3]$; $\Delta T_{1/2} = 0$ °C with a 10-bp duplex determined by FRET-melting). Their binding to G4-DNA is accompanied by an increase of fluorescence intensity by a factor of ~2.6 (**10**: $\lambda_{max} = 448$ nm, **11**: $\lambda_{max} = 450$ nm). The analysis of fluorimetric titration curves revealed an apparent 2:1 ligand–oligonucleotide stoichiometry with an affinity constant per site of around 1.0×10^7 M^{-1} for both derivatives.[19]

Similar behavior has been shown by quercetin (Fig. 30), a flavonoid which has been proved to exhibit antitumor activity. In aqueous solutions quercetin is characterized by two fluorescence bands with maxima at 422 nm (from the locally excited state) and 533 nm (ESIPT band due to the intramolecular proton transfer).

[19] [ligand] 2 μM in Tris–HCl buffer (10 mM, pH 7.2) with 100 mM KCl.

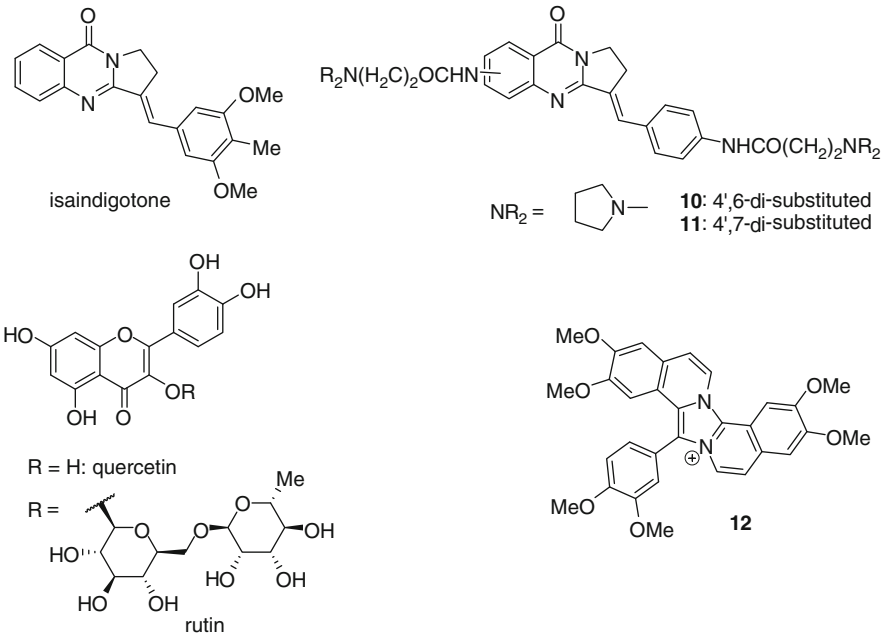
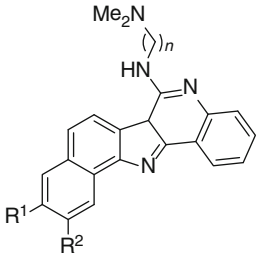


Fig. 30 Natural products binding to quadruplex DNA with increase of their fluorescence

In the presence of tetramolecular G-quadruplexes forming either a monomeric quadruplex unit $(d[T_2AG_3T])_4$ or an end-to-end stacked dimeric quadruplex $(d[T_2AG_3])_4$ the fluorescence intensity of quercetin increases and the ratio between the two bands changes as a function of the quadruplex structure. By analyzing the fluorescence, the absorption, the CD, and the NMR spectra, it was concluded that quercetin interacts differently with the two structures: external π-stacking mode with a monomeric G-quadruplex and groove-binding mode with a dimeric G-quadruplex [132, 133]. A significant fluorescence enhancement was also observed for rutin, a glycoside derivative of quercetin [133].

The fluorenylium derivative **12**, a product of oxidation of papaverine, also exhibits a fluorescence "light-up" behavior. In aqueous solutions **12** is weakly fluorescent due to strong aggregation (H-aggregates in the case of low-salt conditions and J-aggregates at high ionic strength). Upon addition of quadruplex DNA $d[(G_3T_2A)_3G_3]$, fluorescence enhancement by a factor of ~8 is induced, whereas interaction with ds DNA resulted only in a twofold increase of fluorescence [134].

Although moderate, the fluorescence increase of those natural products can be used to evidence and characterize the interaction with quadruplexes. However, it is less likely that it is to be of practical use for labeling or detecting quadruplexes due to (1) the poor magnitude of the effect and (2) the rather poor quadruplex vs duplex selectivity of most of these compounds.



PSI99A: $n = 3$, $R^1 = H$, $R^2 = NO_2$
PSI111: $n = 3$, $R^1 = H$, $R^2 = OMe$
PSI109: $n = 2$, $R^1 = H$, $R^2 = OMe$
SD27: $n = 3$, $R^1 = OMe$, $R^2 = H$

Fig. 31 Structures of the benzoindoloquinolone derivatives

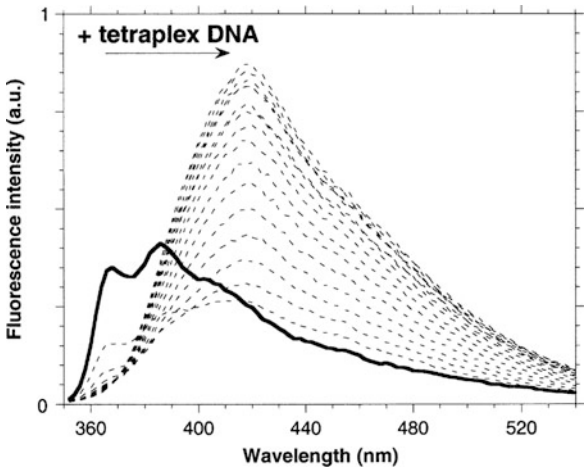


Fig. 32 Fluorescence titration of SD27 with the G-rich strand at 20 °C. The *dashed lines* correspond to SD27 with increasing concentrations of 21AG quadruplex (from 0.1 μM to 2.0 μM). Reprinted with permission from [135]. Copyright 2002 Elsevier

2.6.2 Synthetic Compounds

Benzoindoloquinoline derivatives (Fig. 31) were shown to be moderate to good quadruplex binders ($\Delta T_{1/2}$ = 3–11 °C measured by FRET-melting with F21T) [135]. In this series only SD27 is strongly fluorescent and displays a strong red-shift and a large increase of its emission upon addition of d[$(G_3T_2A)_3G_3$]

(Fig. 32).[20] The structureless emission could originate from formation of excimers and the absence of an isoemissive point indicates the existence of several bound species.

2.7 *Metal Complexes*

Over the past 40 years, metal complexes (mostly cationic) have been extensively studied as DNA binders [136–138]. More recently, this chemical class has been developed for targeting quadruplex nucleic acids [14, 139]. Hereafter, we review the metal-containing species (transition metal complexes, transition metal phthalocyanines, lanthanide cations and their complexes) displaying light-up fluorescence properties upon binding to quadruplex structures.

2.7.1 Diverse Transition Metal Complexes

Platinum(II) Complexes

A family of eight platinum(II) complexes, comprising the aromatic ligands dipyridophenazine (dppz) or C-deprotonated 2-phenylpyridine, were synthesized and studied as luminescent quadruplex-DNA probes [140].

One of the dppz complexes, namely $[Pt^{II}(dppz\text{-}COOH)(N\text{^}C)]CF_3SO_3$ (N^CH = 2-phenylpyridine, **13**, Fig. 33), binds quadruplex structures likely through classical external-quartet stacking with a submicromolar K_d and, more importantly, presents a 293-fold fluorescence intensity increase upon addition of quadruplex DNA ($\lambda_{max} = 512$ nm, human telomeric sequence) at a rather high concentration of the complex (50 μM).[21] This emission was assigned to a $^3[Pt \rightarrow \pi^*(dppz)]$ ^{3}MLCT exited state, which takes place once the complex is isolated from water. Interestingly, the fluorescence enhancement is significantly lower with double-stranded DNA, such as ct DNA, poly(dA-dT)$_2$, and poly (dG-dC)$_2$ (27-, 6-fold, 3-fold, and 6-fold, respectively), which is consistent with the preferential binding to the quadruplex, as shown by competitive dialysis. Thanks to these properties, selective detection of micromolar concentrations of quadruplex DNA has been achieved with **13** in gel electrophoresis (PAGE) using long-wavelength UV irradiation.

The same research group (Ma et al.) developed two other families aimed at down-regulating *c-myc* expression. Hence, nine Schiff base complexes were prepared following a computer-aided structure-based drug design [141]. Emission titrations by a quadruplex structure revealed an eightfold intensity enhancement of the complex **14** ($\lambda_{max} = 652$ nm, Fig. 34).

[20] Buffer conditions: 10 mM sodium cacodylate, 100 mM KCl, pH 7.5. Excitation at 274 nm.

[21] Experiments performed in 10 mM Tris HCl buffer pH 7.5 with 100 mM K^+.

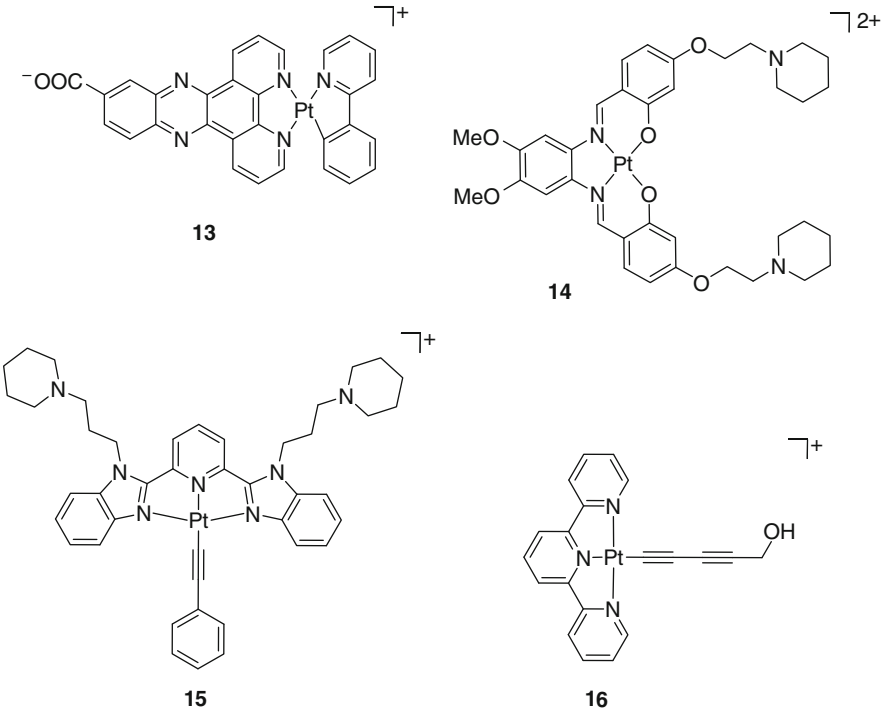


Fig. 33 Platinum(II) complexes

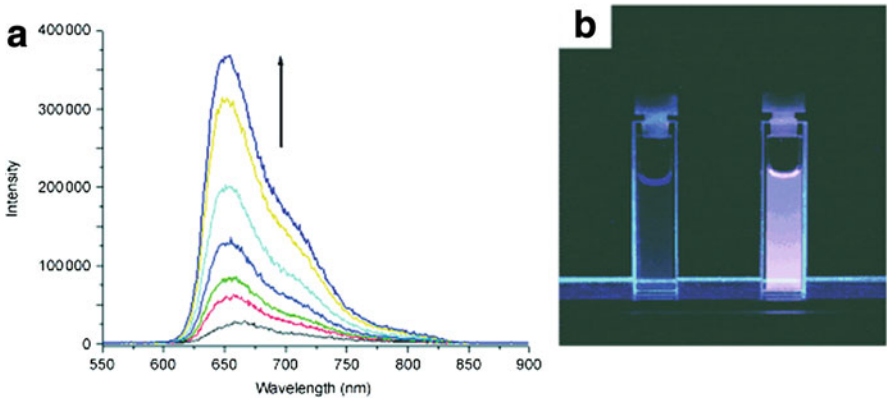


Fig. 34 (**a**) Emission spectra of the complex **14** with increasing concentration of c-myc quadruplex DNA d[$TG_4AG_3TG_4AG_3TG_4A_2G_2$]. (**b**) Photographs of a solution of this complex (50 mM) in absence and presence of c-myc quadruplex DNA. Reprinted with permission from [141]. Copyright 2009 John Wiley and Sons

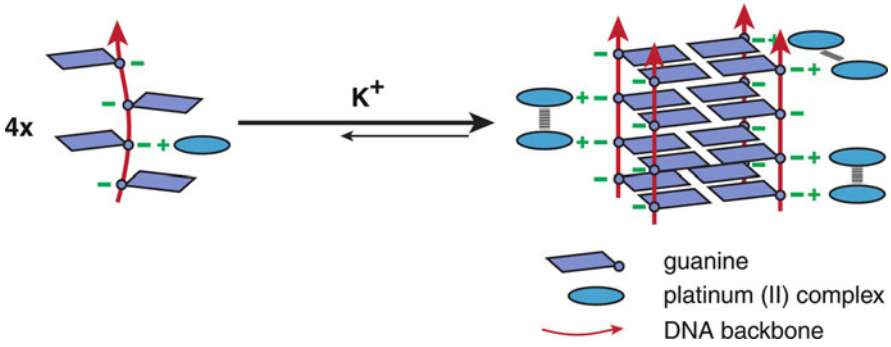


Fig. 35 Schematic cartoon showing the possible self-assembly of **16** via Pt···Pt and π–π interactions induced by G-quadruplex formation upon K^+ ion binding

The second series is composed of seven tridentate bis(benzimidazole)pyridine (bzimpy) or bis(pyrazole)pyridine (dPzPy) platinum(II) complexes. The bzimpy derivative **15** presented in Fig. 33 displays a broad emission band upon binding to quadruplex DNA, with a 38-fold fluorescence intensity increase ($\lambda_{max} = 622$ nm). These complexes bind duplex DNA with at least a tenfold lower affinity than quadruplex DNA, together with a lower fluorescence enhancement (4.4-fold for bzimpy derivative with ct-DNA).

The Pt(II)-terpyridine complex **16** was used to detect G-quadruplex formation via self-assembly of four short guanine-rich strands (Fig. 35) [142]. Thus, short G-rich oligonucleotides were mixed with the terpyridine complex, which resulted in their binding via electrostatic interactions. Upon addition of potassium ions, the oligonucleotides form a quadruplex structure, resulting in a high local concentration of the platinum complex, which in turn leads to emission increase due to the metal–metal-to-ligand charge transfer (^{3}MMLCT) band ($\lambda_{ex} = 550$ nm, $\lambda_{max} = 800$ nm). However, the binding and fluorescence properties of **16** vis-à-vis prefolded intramolecular quadruplex DNA structures have not been studied, so far.

Ruthenium(II) Complexes

Although Ru(II) complexes are mostly nonplanar and feature octahedral or pseudo-octahedral geometry (cf. Fig. 40), there is a growing interest in their use as pharmaceuticals, and notably as DNA binders [143]. Barton's group reported for the first time reversible light-up DNA-binding ruthenium(II) complexes containing dipyrido[3,2-*a*:2′,3′-*c*]phenazine (dppz) and bipyridine (bpy) or phenanthroline (phen) ligands [144, 145]. These compounds are almost nonemissive in aqueous

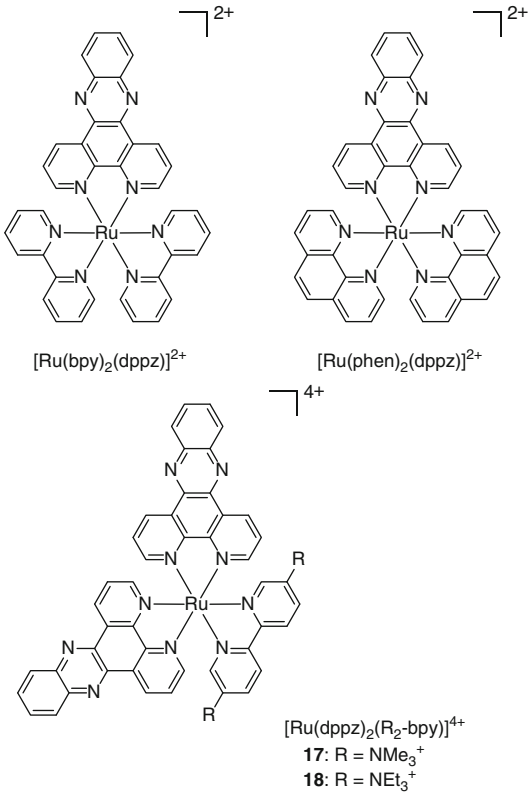


Fig. 36 Mononuclear ruthenium(II) complexes. Counter-ions are hexafluorophosphates

buffers but a ^{3}MLCT excited state fluorescence emission occurs upon DNA intercalation.

Although many studies with duplex DNA have been published since then, interaction of $[Ru(phen)_2(dppz)]^{2+}$ and $[Ru(bpy)_2(dppz)]^{2+}$ with quadruplex DNA was only recently examined (Fig. 36) [146, 147]. $[Ru(phen)_2(dppz)]^{2+}$ binds to the human telomeric sequence in potassium conditions more tightly than in sodium conditions, or to the i-motif (one order of magnitude difference in the latter case).[22] Addition of quadruplex to the complex results in a fluorescence intensity increase ($\lambda_{max} \approx 625$ nm), around five times larger with the quadruplex structure as compared to the i-motif, likely due to a better π-overlapping and a better isolation from the solvent. Analogous results were obtained with $[Ru(bpy)_2(dppz)]^{2+}$. With the human telomeric sequence, both the fluorescence enhancement (fourfold) and the binding affinity (one order of magnitude) of the complex were larger in potassium-rich buffer than in sodium-rich buffer. In either case, the quadruplex-forming sequence was preferred over the i-motif structure.

[22] Experiments performed at pH 5.5.

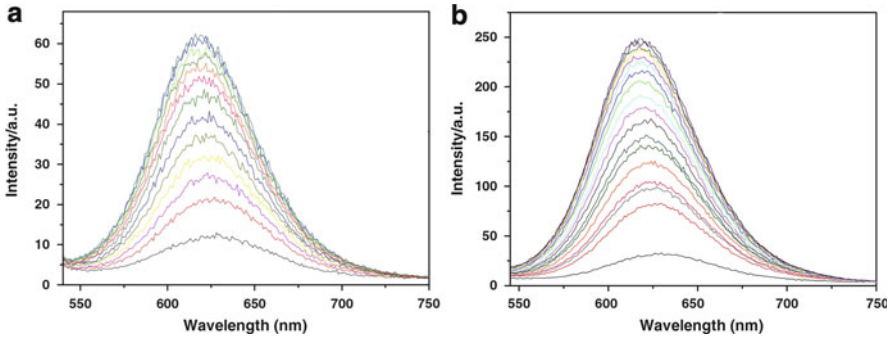


Fig. 37 Emission spectra of complexes $[Ru(dppz)_2(R_2\text{-bpy})]^{2+}$ (**a**: **17**, **b**: **18**; 2 μM) in the presence of telomeric quadruplex DNA ($AG_3(T_2AG_3)_3$, 0–5 μM). Reprinted with permission from [148]. Copyright 2011 Elsevier

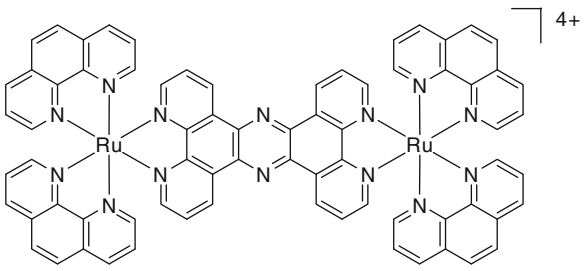
Two analogous complexes, but containing two dppz ligands, were prepared [148]. The third ligand is a bpy derivative, functionalized by either trimethylammonium (**17**) or triethylammonium terminal groups (**18**) (Fig. 36). Stabilization of human telomeric sequence (determined by UV-melting at a 1:1 ratio) induced by both complexes was higher than with $[Ru(bpy)_2(dppz)]^{2+}$ [146]. The derivative **17** binds more tightly than **18** (9.4°C vs 7.0 °C), with an apparent 1:1 stoichiometry. Both complexes exhibit a marked fluorescence intensity increase upon addition of the human telomeric quadruplex, but to a larger extent for **18** than for **19** (7.6-fold and 4.8-fold, respectively; $\lambda_{max} = 630$ nm; Fig. 37).[23]

The Thomas' group investigated the spectroscopic properties of two dinuclear Ru(II) complexes **19** and **20** containing a tppz (tetrapyrido[3,2-*a*:2′,3′-c:3″, 2″-*h*:2″,3″-*j*]phenazine) and bipyridine or phenanthroline ligands, upon duplex or quadruplex DNA binding (Fig. 38) [149]. Interaction of these two compounds with double-stranded ct DNA results in a large fluorescence enhancement (>60-fold, $\lambda_{max} = 658$ and 637 nm for the phen and bpy complexes, respectively).[24] The light-up effect has been rationalized by analogy with the aforementioned dppz complexes via a computational study [150]. Addition of human telomeric quadruplex DNA in high ionic strength conditions[25] induced a fluorescence intensity enhancement of the metal complexes, ca. 2.5 times higher than with duplex DNA (150-fold), together with a substantial hypsochromic shift as compared to ds DNA ($\lambda_{max} = 631$ and 605 nm for the phen and bpy complexes, respectively) and longer emission lifetimes (129 and 123 ns for G4-DNA vs 84 and 92 ns for ds DNA). These data were explained by a better isolation of the complexes from the aqueous medium and

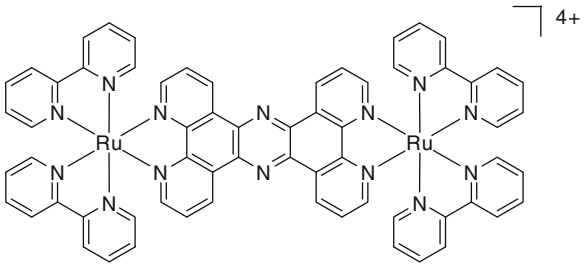
[23] Experiments performed at pH 7.0, in 10 mM potassium phosphate buffer, with 100 mM KCl and 1 mM K_2EDTA.

[24] Experiments performed at pH 7.0, in 5 mM Tris buffer, 25 mM NaCl.

[25] Experiments performed at pH 7.0 in 10 mM potassium phosphate buffer with 200 mM KCl and 1 mM K_2EDTA.



[(phen)$_2$Ru(tppz)Ru(phen)$_2$]$^{4+}$(**19**)



[(bpy)$_2$Ru(tppz)Ru(bpy)$_2$]$^{4+}$ (**20**)

Fig. 38 Dinuclear ruthenium (II) complexes

a greater aromatic overlap with the DNA bases through a 1:1 end-stacking mode. It was also proposed, on the basis of calorimetric studies, that stacking of **19** to a quadruplex structure is more favorable than intercalation into duplex DNA (5 kJ mol^{-1} difference). However, the proposed end-stacking mode is still questionable considering the large size and complex 3D topology of these dinuclear octahedral Ru(II) complexes.

Other quadruplex-forming sequences (TBA, Pu27 from *c-myc*, the *Oxytricha* telomeric sequence) were studied with the phenanthroline complex **19** [206]. The intensity of the ^{3}MLCT emission band and maximum wavelength are both strongly affected by the quadruplex structure. Indeed, fluorescence enhancement ranges from threefold (with TBA) to 138-fold (with the human telomeric sequence), with the maximum emission wavelength varying accordingly (the higher the enhancement, the larger the hypsochromic shift) from 635 nm (human telomere) to 670 nm (TBA). However, the affinities (associated with the 1:1 end-stacking) and fluorescence enhancements are not directly correlated, since larger emission enhancement are also dependent on isolation from the solvent [145]. The authors have explained these differences by the nature and the length of the quadruplex loops, since external loops containing at least three bases result in higher fluorescence intensities of the bound ligand than small lateral loops. Results obtained with complex **20** are analogous.

Finally, it should be noted that **19** was used as a multifunctional biological imaging agent for staining nuclear DNA of living eukaryotic and prokaryotic

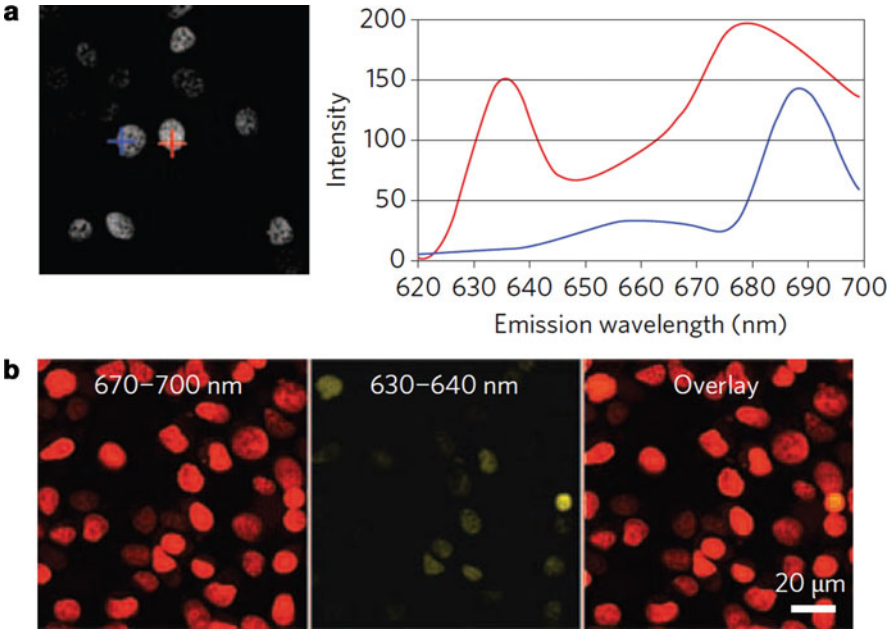


Fig. 39 (**a**) Multiple-emission profile of **19** in live MCF-7 cells, in two separate cellular regions (*red* and *blue*); (**b**) Confocal microscopy images with separate detection channels: 670–700 nm (*red*) and 630–640 nm (*yellow*). Reprinted with permission from [151]. Copyright 2009 Nature Publishing Group

cells. However, exceptionally large concentrations of compound are required (i.e., 500 µM) which is attributed to an inefficient nonendocytotic cellular uptake [151]. More interestingly, a dual emission peaking at 630 and 680 nm has been observed from the nuclei stained by the Ru complex, which was interpreted as concomitant labeling of duplex DNA and alternative DNA structures (potentially quadruplexes). The blue-shifted band (630 nm) is observed only when the probe is incubated with the cells before fixation by formaldehyde, thereby suggesting that these alternative DNA structures may be induced by probe binding (Fig. 39).

Another rigid, dinuclear but unsymmetrical ruthenium(II) complex **21**, containing the obip [2-(2-pyridyl)imidazo[4,5-*f*][1,10]phenanthroline] and bpy ligands, has been studied (Fig. 40) [152]. Addition of the human telomeric quadruplex in sodium conditions resulted in a moderate fluorescence enhancement (1.5-fold, $\lambda_{ex} = 460$ nm, $\lambda_{max} \approx 620$ nm), although higher than with calf-thymus DNA.[26] Concurrently a tenfold preference for the quadruplex over duplex structure was found by fluorimetric titration (1:1 electrostatic external and/or end-stacking model).

[26] Experiments performed in 100 mM NaCl, 10 mM NaH_2PO_4/Na_2HPO_4, 1 mM Na_2EDTA, pH 7.0.

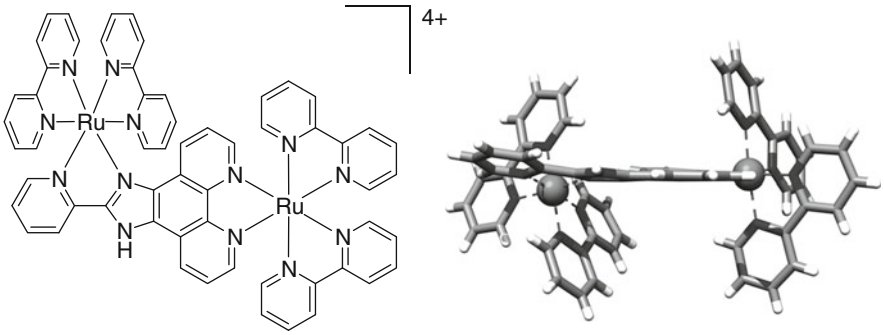


Fig. 40 Structure of $Ru_2(obip)(bpy)_4$. *Right*: crystal structure by X-ray diffraction (C: *gray*, H: *white*, N: *blue*, Ru: *turquoise*)

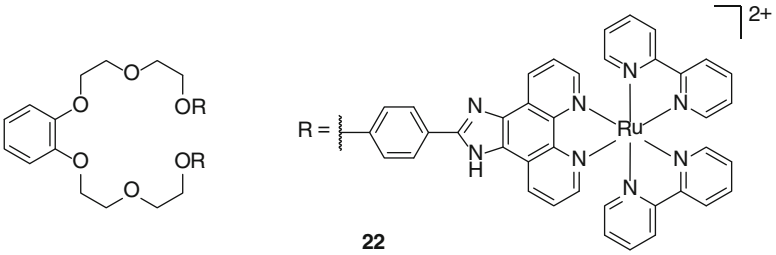


Fig. 41 Structure of the flexible dinuclear ruthenium complex. The podand linker has been depicted in a pseudocyclic conformation, but remains flexible in the absence of potassium cations

A dinuclear ruthenium(II) complex **22** was designed using a partially flexible acyclic chain instead of a rigid ditopic ligand [153]. Both Ru^{2+} cations are complexed by two bpy and one hpip (2-phenyl-1H-imidazo[4,5-*f*][1,10] phenanthroline) ligands (Fig. 41). It is supposed that upon addition of potassium cations, the podand linker adopts a pseudocyclic conformation in order to bind potassium, resulting in a fluorescence intensity decrease, likely due to ruthenium self-quenching when both metal centers come into close proximity. When various quadruplex structures (human telomere, Bcl2, Pu18, c-kit, Vegf) were added, a fivefold increase of fluorescence intensity was observed, presumably due to the opening of the pseudocyclic system ($\lambda_{ex} = 468$ nm, $\lambda_{max} \approx 600$ nm).[27] Conversely, only a slight increase was observed with ds DNA and even a decrease with ss DNA.

However, strong background fluorescence remains caused by the unbound ruthenium species. This was eliminated using the heavy atom effect by the addition

[27] Experiments performed in 10 mM Tris–HCl buffer, pH 7.4, with 100 mM KCl.

of 50 mM iodide ions. The authors assumed that iodide anions cannot access the bound complexes due to electronic repulsion of DNA phosphates, and fluorescence was indeed recovered upon addition of quadruplex structures, with a 15-fold fluorescence selectivity for c-kit against ds DNA. Using this system, a high sensitivity was obtained, enabling the distinction between quadruplex and duplex DNA at submicromolar concentrations of both **22** and DNA. Thus, as little as 50 nM of quadruplex DNA could be detected in the presence of 1 μM of ds DNA. As found with various other probes described previously, fluorescence enhancement is not only influenced by the binding affinity but also by the binding mode as this complex binds quadruplex structures with only a slight preference over duplex DNA as shown by SPR measurements. A double external stacking model was proposed for this compound, where each ruthenium unit binds one of the external G-quartets.

2.7.2 Transition-Metal Phthalocyanines

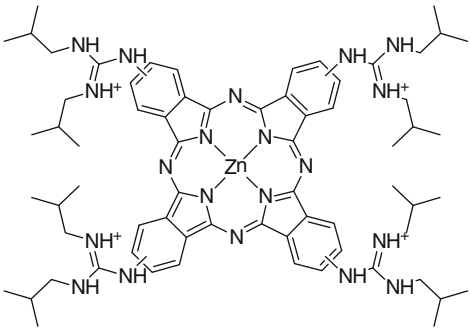
Since the discovery of phthalocyanines more than 80 years ago, this class of macrocyclic compounds, and the corresponding coordination complexes, have been used extensively as fluorescent dyes. First studies of the quadruplex-binding properties of pyridinium-porphyrazine and ammonium-porphyrazine derivatives revealed an enhanced selectivity compared to the well-known porphyrin TMPyP4 (see Sect. 2.4) [154, 155].

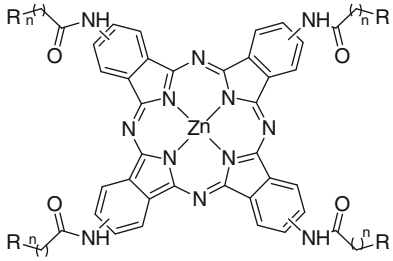
Luedtke's group developed zinc complexes of guanidinium derivatives of phthalocyanine to improve both the cellular uptake and the nucleic acid affinity [156]. Zn-DIGP (Fig. 42) displays a large fluorescence enhancement (200-fold, $\lambda_{ex} = 620$ nm, $\lambda_{max} = 705$ nm) upon addition of saturating amounts of nucleic acids.[28] The resulting fluorescence quantum yields are rather low ($\Phi_F = 0.06$) but counterbalanced by very large molar extinction coefficients ($\varepsilon = 30{,}000$–$130{,}000$ $cm^{-1}\ M^{-1}$) that result in a strong brightness ($\Phi_F \times \varepsilon$), the relevant figure of merit for imaging. Remarkably high binding constants were extracted from fluorimetric titration with various quadruplex structures. In particular, with the c-myc sequence a nanomolar K_d was found ($K_d = 2 \times 10^{-9}$ M per site with a 2:1 stoichiometry), with a preference of one order of magnitude over unfolded G-rich ss DNA, and 1,000-fold, 100-fold, and 5,000-fold selectivity over C-rich unfolded ss DNA, tRNA, and calf thymus DNA, respectively.

Fixed and living cell microscopy (wide-field or confocal) revealed a successful internalization of Zn-DIGP, mostly probing trafficking vesicles and perinuclear organelles, with no staining of duplex DNA as seen by Hoechst 33342 co-staining experiments (Fig. 43).

A K^+ detection system has been designed, using Zn-DIGP and the parallel quadruplex forming sequence c-myc [157]. This system relies on the promotion

[28] Experiments performed in 50 mM Tris–HCl, pH 7.4, with 150 mM KCl and 0.5 mM EDTA.





APc

	n	R
1	1	NH_3^+
2	2	NH_3^+
3	3	NH_3^+
4	4	NH_3^+
5	2	$N^+(CH_2)_4$
6	2	COO^-

Fig. 42 Structure of the guanidinium-modified and amido-modified phthalocyanines

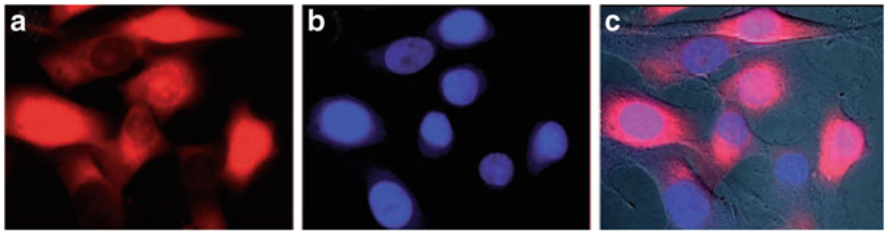


Fig. 43 Fixed SK-Mel-28 cells stained with 3 μM Zn-DIGP (**a**; $\lambda_{ex} = 620$ nm, $\lambda_{max} = 700$ nm) and 8 μM Hoechst 33342 (**b**; $\lambda_{ex} = 360$ nm, $\lambda_{max} = 470$ nm) and overlay (**c**). Adapted with permission from [156]. Copyright 2009 John Wiley and Sons

of the c-myc folding by K^+, followed by an easily observable fluorescence intensity increase of Zn-DIGP, once bound to the quadruplex structure. A 0.8 μM detection limit of K^+ was determined. The absence of promotion of c-myc quadruplex structure by various other cations (NH_4^+, Na^+, Ca^{2+}, Mg^{2+}, Zn^{2+}, Fe^{3+}, and Cu^{2+}) allows a specific detection of potassium. Hence, 40 μM of K^+ could be detected in the presence of a 3,500-fold excess of Na^+ ions.

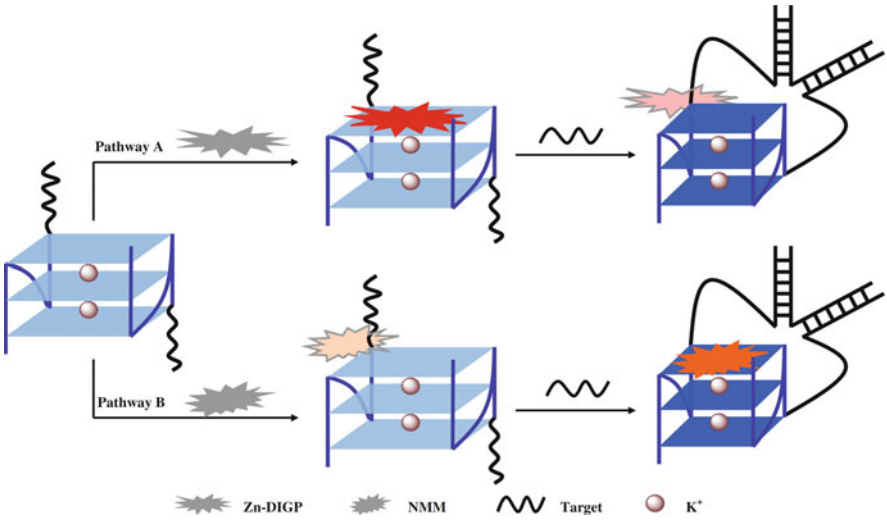


Fig. 44 Schematic illustration of DNA sensors based on turn-off (*pathway A*) and turn-on (*pathway B*) fluorescence changes, utilizing split G-quadruplex probes. In the presence of K^+, a split c-myc (*blue*) forms an associated G-quadruplex-fluorescent dye complex. Sequence-specific DNA hybridization results in reduced fluorescence from Zn-DIGP (*pathway A*) or increased fluorescence from NMM (*pathway B*). Reprinted with permission from [158]. Copyright 2011 Springer

Recently, Zn-DIGP was used jointly with the *N*-methylmesoporphyrin IX (NMM) in a nucleic acid detection assay [158]. Briefly, the c-myc sequence was separated into two fragments, and flanking segments complementary to a target DNA sequence were added (Fig. 44). In the presence of potassium cations, the two sequences readily form a dimeric quadruplex structure, which is further stabilized by the bound fluorescent probes. Upon addition of the target sequence, base-pairing interactions with both flanking sequences induce the formation of a quadruplex-duplex three-way junction, which modifies the environment and consequently the fluorescence of the probe. This fluorescence modification (increase for NMM, decrease for Zn-DIGP) allows the detection of the target sequence.

A second class of acylamino-substituted phthalocyanine derivatives, in which systematic variations of the length of the alkyl chain and the nature of the terminal function were performed (APc, Fig. 42), was studied by the same group [159]. The influence of these variations on the fluorescence light-up and quadruplex binding properties was subsequently examined. Large enhancements were observed with APc 2–5 (100–400-fold), but not with APc 1 and 6, in the presence of the human telomeric quadruplex ($\lambda_{ex} = 630$ nm, $\lambda_{max} = 705$ nm) (see footnote 28). APc 2 displayed the best performances (400-fold increase, $\Phi_F = 0.05$, $\varepsilon = 30\,000$–$150{,}000\ cm^{-1}\ M^{-1}$). Tenfold lower binding constants were determined for APc 2–5 as compared to Zn-DIGP, but the high selectivity against tRNA and calf thymus DNA is preserved (50-fold and 500-fold, respectively). Large fluorescence enhancements were also

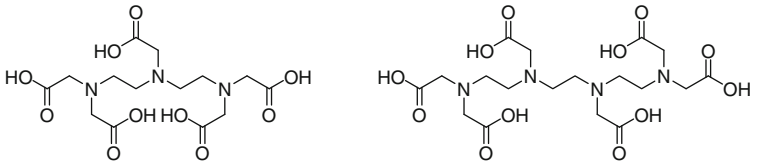


Fig. 45 Structure of DTPA and TTHA ligands

observed with c-myc and c-kit2 structures (200–800-fold), and the following relative binding affinity sorting was suggested: c-myc > c-kit2 > human telomeric quadruplex. It was hypothesized that APc 1 forms J-aggregates unable to bind quadruplex DNA whereas APc 2–5, which contain more methylene units and hence increased pK_a values, form nonfluorescent H-aggregates. Fluorescence enhancement is thus the result of dissociation of the APc aggregates into quadruplex-binding monomers.

2.7.3 Lanthanides

Lanthanide complexes are extensively used as emitters for luminescent resonance energy transfer and as contrast agents in magnetic resonance imaging, but have also already been used to detect duplex DNA via fluorescence enhancement [160–163]. Terbium(III) and europium(III) ions have well-suited luminescence properties for nucleic acid probing, such as narrow emission band widths and large Stokes shifts (see Fig. 47 for an example), but Tb^{3+} emitting level fits better the triplet state of nucleic acid bases [163, 165] (for a review on the "antenna effect" see [166]).

Terbium(III) emission in aqueous medium is very weak because of the solvent quenching and a low intrinsic molar absorptivity [165]. Upon binding of $TbCl_3$ to quadruplex DNA, energy is transferred via a FRET mechanism from the bases (donor, $\lambda_{ex} = 290$ nm) to the terbium ion (acceptor) inducing a $^5D_4$ long-lived excited state emission ($\lambda_{max} = 545$ nm), which results in a ~25-fold luminescence increase of Tb^{3+} in the presence of human telomeric DNA [167]. Interaction of Tb^{3+} ions with quadruplex DNA is believed to occur through coordination of nitrogen and oxygen atoms of bases and phosphate groups, with a micromolar affinity and a 6:1 stoichiometry.[29] Putative binding sites are the quadruplex central channel and TTA loops. Cation-displacement experiments were performed with potassium, sodium, and lithium (in that ranking of efficiency), resulting in quenching of terbium fluorescence, which may evidence the presence of terbium in the quadruplex central channel (but also increase the ionic strength that may affect the fluorescence).

[29] Experiments performed in 10 mM Tris–HCl buffer, pH 7.0, without other cations.

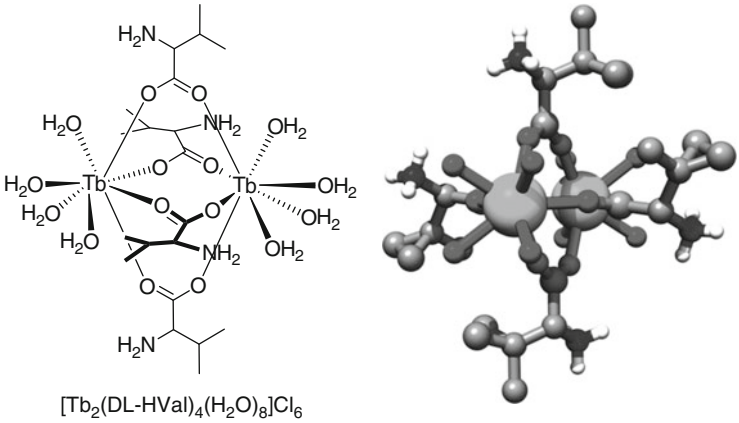


Fig. 46 Terbium-amino acid complex structure and solid-state structure (from X-ray diffraction; C: *gray*, H: *white*, N: *blue*, Tb: *green*)

Förster energy transfer was used with reverse conditions, i.e., terbium(III) ions (chelated by DTPA or TTHA, Fig. 45) as donors and quadruplex DNA as acceptor [168]. The change in fluorescence lifetime ($\lambda_{ex} = 488$ nm) was monitored in the presence of small amounts of quadruplex DNA (20–800 nM), allowing to calculate the amount of transferred energy. Hence, this amount increased with quadruplex DNA concentration (single-stranded 22-mer and double-stranded 12-mer).

Nonchelated europium(III) was used in the same study to detect low quadruplex DNA concentrations (70% and 25% fluorescence enhancements with 800 nM of quadruplex or duplex calf thymus DNA, respectively; $\lambda_{max} = 453$ nm).

Finally, interactions between two terbium-amino acid (DL-cysteine or DL-valine) complexes (**23**; Fig. 46) and quadruplex or i-motif DNA were studied in 2006 [169]. Both complexes destabilize slightly the human telomeric quadruplex ($\Delta T_m = -3.0$ to -4.0 °C, UV-melting), presumably through coordination binding of terbium to N7 of purines and exocyclic C2 of pyrimidines [164, 170]. Binding constants with quadruplex and i-motif DNA were analogous and rather low (ca. 4×10^4 M^{-1}). A 253-fold and 88-fold fluorescence increase (for cysteine and valine complexes, respectively) were observed in the presence of quadruplex DNA, as compared with a 32-fold increase for the control complex $TbCl_3$ ($\lambda_{ex} = 260$ nm, $\lambda_{max} = 488$, 534, 584 nm; Fig. 47).[30]

Interestingly, the fluorescence increase in the presence of i-motif was significantly lower (around threefold), despite a comparable affinity of the complex for both structures. The authors suggested that the more efficient energy transfer between guanine bases and terbium may account for these results [163, 171].

[30] Experiments performed in 50 mM Tris-HCl buffer, pH 7.1, with 100 mM NaCl.

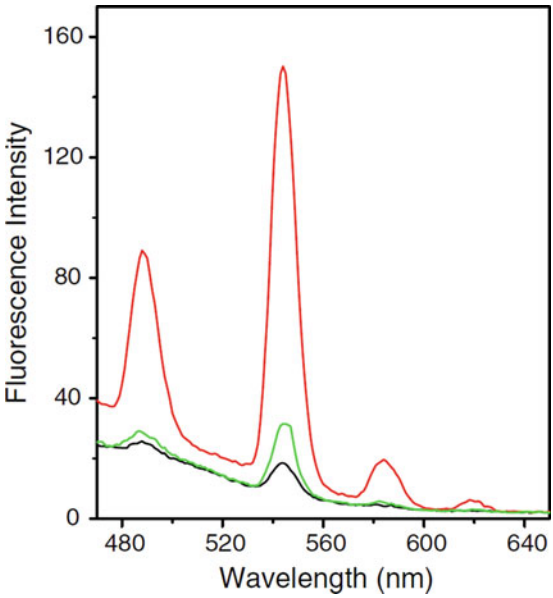


Fig. 47 Fluorescence spectra of $[Tb_2(\text{DL-HVal})_4(H_2O)_8]Cl_6$ in absence (*black*) or presence of quadruplex (*red*) and i-motif (*green*). Reprinted with permission from [164]. Copyright 2006 Elsevier

3 "Light-Off" Probes

Beyond the ligands whose binding to quadruplex DNA is accompanied by an increase or shift of the fluorescence signal, a number of small molecules – usually heterocyclic derivatives – have been described which are intrinsically fluorescent in the absence of DNA, but whose fluorescence intensity decreases to a smaller or larger extent upon binding to the quadruplex target. This fluorescence quenching phenomenon is in most cases due to an electron transfer reaction between the excited fluorophore and DNA, which is exergonic when the reduction potential of the excited dye (E_{red}^{*}) is larger than the oxidation potential of the nucleic bases. As guanine has the lowest oxidation potential among the DNA bases (1.02 V vs NHE for guanosine and GMP at pH 7.4) [172], it represents the base which is most easily oxidized, and the high density of guanine residues in a quadruplex DNA makes the electron-transfer reaction with a fluorophore even more favorable. Moreover, it was reported that the oxidative damage upon photooxidation of quadruplex DNA is localized on the guanines of the external G-tetrad [173], which gives some evidence that their oxidation potential is lower than the one of guanine bases in duplex DNA. Since the external G-tetrad represents the binding site for a vast majority of quadruplex ligands, it is not unexpected that binding of many fluorescent ligands is accompanied by an efficient quenching of their fluorescence. Nonetheless, it should be mentioned that quenching of fluorescence is not as desired in analytical or

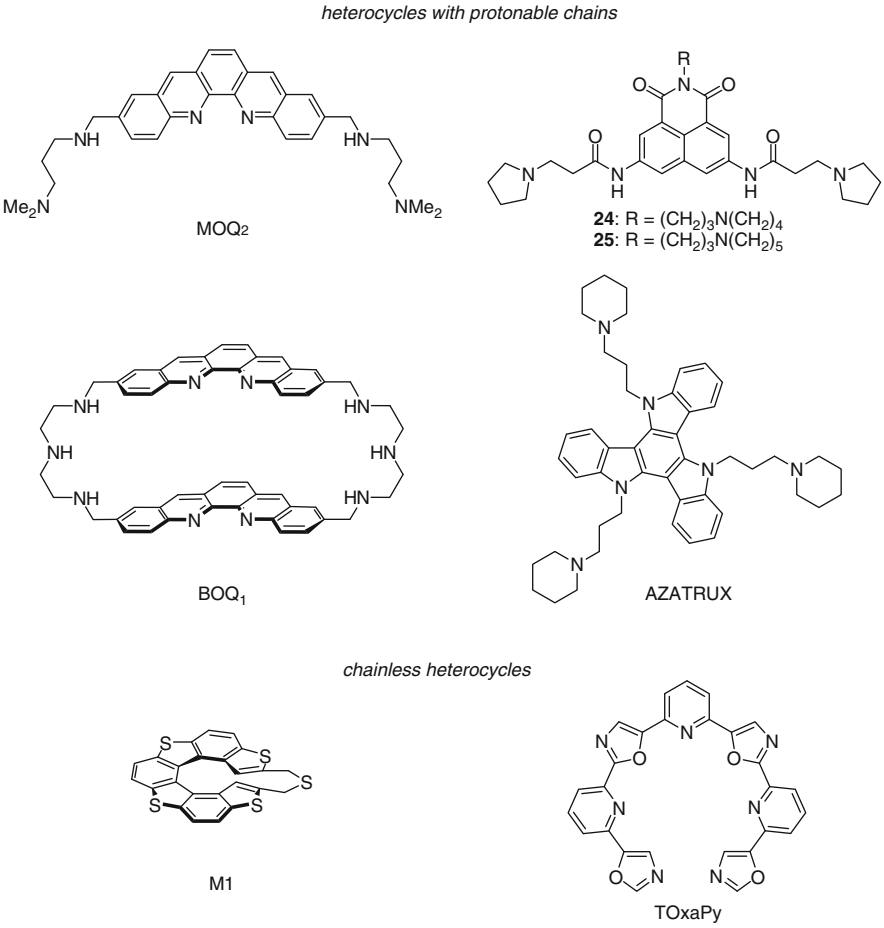


Fig. 48 Heterocyclic ligands with fluorescence-quenching behavior upon binding to quadruplex DNA

biomedical applications as fluorescence shift or fluorescence enhancement, since quenching might be also due to other species present in samples, such as molecules favoring the intersystem crossing (amines, carbonyl derivatives), inorganic quenchers (I^-, $[Fe(CN)_6]^{3-}$), or other processes such as ligand aggregation. Still, the fluorescence quenching phenomenon is very useful when additional information about the binding event has to be obtained (binding constant, stoichiometry) and, if the ligand would be selectively quenched by quadruplex DNA and not by the duplex form, sensitive detection of the quadruplex form may be envisaged.

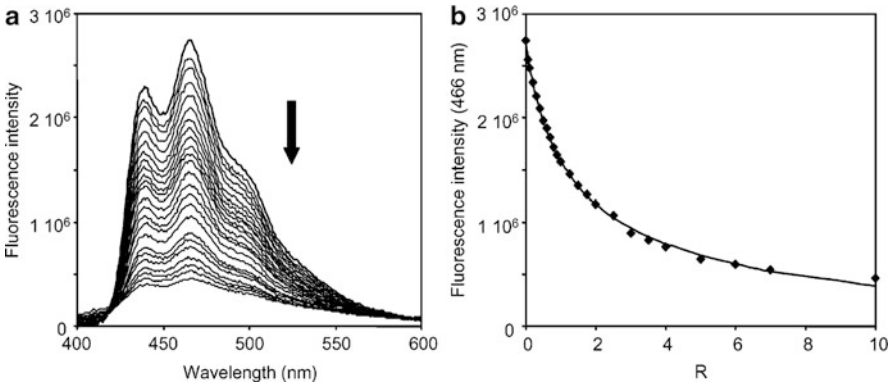


Fig. 49 (**a**) Fluorescence spectra of **MOQ_2** (0.1 μM; $\lambda_{ex} = 360$ nm) recorded at increasing concentration of oligonucleotide 22AG. (**b**) Experimental (*diamonds*) and calculated (*solid line*) titration curves obtained for the 1:2 stoichiometry. R = DNA/ligand molar ratio. Reprinted with permission from [174]. Copyright 2003 American Chemical Society

3.1 Heterocyclic Ligands with a Neutral Core

Most planar aromatic heterocyclic compounds exhibit moderate to strong fluorescence if they are not aggregated in aqueous solutions. Some representative examples of quadruplex DNA ligands whose fluorescence quenching behavior upon binding to the quadruplex target was well-documented are given in Fig. 48. Among them the quinacridine derivatives, such as MOQ_2, were ones of the first identified quadruplex-stabilizing telomerase inhibitors. These heterocycles bearing polyammonium side chains show a bright blue fluorescence in water (e.g., $\lambda_{max} = 462$ nm, $\Phi_F = 0.10$ for MOQ_2), which is quenched nearly completely upon binding to the telomeric G-quadruplex 22AG (Fig. 49). The analysis of fluorimetric titration curves performed at increasing concentration of 22AG revealed a binding constant of 3.4×10^6 M^{-1} and a 2:1 stoichiometry [175].[31] The bis-quinacridine macrocycle BOQ_1 was shown to bind 22AG with a higher affinity than the monomeric quinacridine and a tenfold selectivity with regard to duplex DNA ($K_a = 1 \times 10^7$ M^{-1} with 22AG; $K_a = 1.2 \times 10^6$ M^{-1} with a 20-bp duplex, 2:1 stoichiometry in both cases, as determined by SPR) (see footnote 31). The fluorescence of BOQ_1 in the free state is reduced with respect to the monomeric quinacridine due to the intramolecular stacking effect ($\lambda_{max} = 466$ nm, $\Phi_F = 0.03$), and binding to the quadruplex leads to further quenching of the fluorescence down to ≈50% of the original value. In contrast, addition of duplex DNA leads to only a slight decrease of the fluorescence consistent with the selectivity inferred from SPR data [174].

[31] In 10 mM HEPES, 200 mM KCl, 150 mM NaCl, pH 7.4.

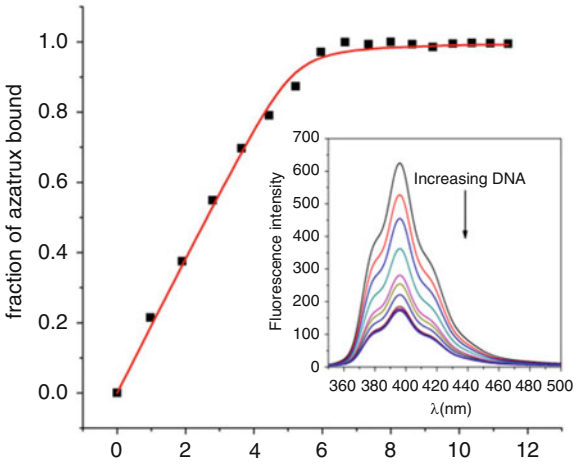


Fig. 50 Fluorescence emission spectra of AZATRUX (8 μM) in the presence of successive additions of telomeric G4-DNA (*inset*) and plot of the fraction of bound AZATRUX vs DNA concentration. Reprinted with permission from [179]. Copyright 2011 Elsevier

The trisubstituted naphthalimides **24** and **25** were selected as potent G-quadruplex ligands from a series of similar derivatives by means of fluorimetric titrations. Their interaction with 22AG leads to quenching of naphthalimide fluorescence, which was useful for determination of the binding stoichiometry and the binding constants (2:1 complexes, $K = 3.0 \times 10^6$ and 2.4×10^6 M^{-1} for **24** and **25**, respectively) [176].[32]

AZATRUX is a heterocycle with a large C_{3h}-symmetric aromatic core, designed to stack over the G4 tetrad, and three aminoalkyl chains [177, 178]. In solution AZATRUX shows a maximum of fluorescence at 394 nm ($\lambda_{ex} = 320$ nm) that is strongly decreased upon interaction with the telomeric sequences forming dimeric and monomeric quadruplexes d[$(TTAG_3)_8TT$], d[$(AG_3TT)_4$]. At saturation the fluorescence is completely quenched in the presence of the dimeric quadruplex, whereas a residual fluorescence is detected in the presence of the monomeric quadruplex. Binding curves extrapolated from the titrations show binding constants in the (0.5–1) $\times$ 10^6 M^{-1} range (Fig. 50).

The neutral chiral cyclic helicene M1 was designed as a ligand for higher-order quadruplexes, such as quadruplex dimers. It is a strongly fluorescent compound ($\lambda_{em} \approx 475$ nm) whose fluorescence is efficiently quenched (by ≈80%) upon binding to a quadruplex dimer in which the two quadruplex units are separated by a short TTA loop. Interestingly, the fluorescence of the corresponding *P*-isomer is not affected in the same conditions, which indicates that only the *M*-isomer M1 binds to the quadruplex dimer. This unique binding mode was confirmed by molecular modeling calculations [180, 181].

[32] In 20 mM potassium phosphate, 70 mM KCl, 0.1 mM EDTA, pH 7.0.

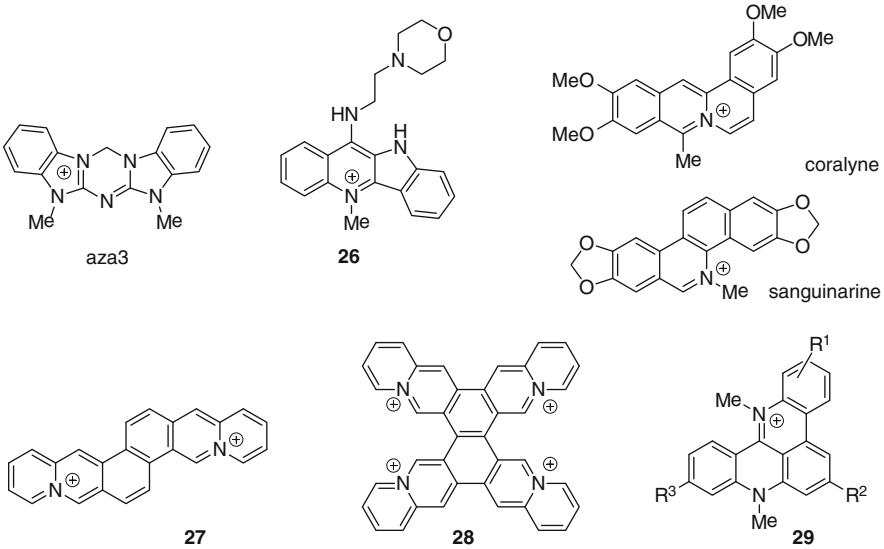


Fig. 51 Cationic ligands whose fluorescence quenching upon binding to quadruplex DNA was documented

Finally, the acyclic polyheteroaromatic pyridyloxazole ligand TOxaPy represents a neutral quadruplex-DNA ligand with an exceptionally high selectivity for human telomeric quadruplex in Na^+-rich conditions, presumably due to its quadruplex groove-binding mode. TOxaPy is characterized by an intense blue fluorescence in water ($\lambda_{em} = 390$ nm, $\Phi_F = 0.50$) which is strongly quenched upon addition of 22AG in Na^+ buffer but much more moderately in the K^+-rich one, which is consistent with the melting data showing unprecedented selectivity for the Na^+ form of the telomeric quadruplex ($\Delta T_m = 10.8$ and <1.0 °C with F21T in Na^+ and K^+, respectively). Analysis of fluorimetric titration data gave a binding constant value of $K = 5 \times 10^6$ M^{-1} with a 1:1 binding stoichiometry [31].[33]

3.2 *Heterocyclic Ligands with a Cationic Core*

Among other reported G-quadruplex ligands, the azacyanine aza3 (Fig. 51) was developed as a readily accessible, efficient and selective ligand for telomeric quadruplex DNA (association constant of 1.3×10^6 M^{-1} with sequence d

[33] In 10 mM $LiAsMe_2O_2$, 100 mM NaCl, pH 7.2.

[$T_2G_3(T_2AG_3)_3A$] vs 9.2×10^3 M^{-1} for calf thymus DNA from fluorimetric titrations, i.e., more than a 100-fold selectivity, confirmed also by SPR) [182].[34] Binding of aza3 to G-quadruplex leads to about twofold quenching of its blue fluorescence ($\lambda_{max} \approx 375$ nm). It may be speculated that this high selectivity with respect to duplex DNA could lead to development of a fluorescence quenching assay for detection of quadruplex DNA.

The cryptolepine derivative **26** is a member of a series of quadruplex-selective quindoline derivatives [56]. Its interaction with the human telomeric G-quadruplex leads to fluorescence quenching accompanied by a hypsochromic shift of the emission maximum (from 472 nm to 461 nm) [183]. Similar quenching behavior was observed for isoquinoline alkaloids coralyne ($\lambda_{max} = 472$ nm), which is also known for its interaction with triplex DNA, and sanguinarine [126]. Synthetic diazoniadibenzochryzene **27** and tetraazoniapentaphenopentaphene **28** belong, like coralyne, to the family of annelated quinolizinium derivatives and possess high fluorescence quantum yields in the absence of DNA (**27**: $\lambda_{em} = 456$ and 481 nm, $\Phi_F = 0.47$; **28**: $\lambda_{em} = 511$ nm, $\Phi_F = 0.14$). Interaction of these highly charged organic cations with human telomeric quadruplex results essentially in complete quenching of their fluorescence, which was used for the calculation of binding constants ($K_a = 1.6 \times 10^6$ and 9.0×10^5 M^{-1} for **27** and **28**, respectively).[35] Interestingly, both **27** and **28** have relatively high quadruplex-vs-duplex selectivity, as demonstrated by thermal denaturation and FID studies [184].

Finally, the *N*-quaternized quino[4,3,2-*kl*]acridinium salts **29** were developed as quadruplex binders with associated potential antitumor activity. Of note, the most well studied member of this series is the well-known quadruplex binder RHPS4 ($R^1 = R^3 = F$, $R^2 = H$), whose fluorescence properties have not been studied in detail. The fluorimetric titrations with 22AG evidenced moderate quenching effect and were used for calculating the quadruplex-binding constants, which were in the range from 0.9×10^5 M^{-1} (**29a**, $R^1 = R^2 = R^3 = H$) to 1.1×10^6 M^{-1} (**29b**, R^1 = 2-NHCO*t*Bu, $R^2 = R^3 = H$) (see footnote 35). In some cases, high selectivity with respect to duplex DNA (with a binding constant ratio of up to 15) was observed in the fluorimetric titrations. However, the quenching was not complete even for the most quadruplex-affine member of the series ($F/F_0 = 0.85$) [185]. These results demonstrate that the efficiency of the fluorescence quenching (represented by the F/F_0 value in the presence of excess quadruplex DNA) and the quadruplex-binding affinity are, in general, not interconnected values, and data analysis should always be performed carefully in order to obtain correct values of binding constants.

[34] In 25 mM K_2HPO_4/KH_2PO_4, 70 mM KCl, pH 7.

[35] In 6 mM Na_2HPO_4, 2 mM NaH_2PO_4, 185 mM NaCl, 0.1 mM EDTA, pH 7.0.

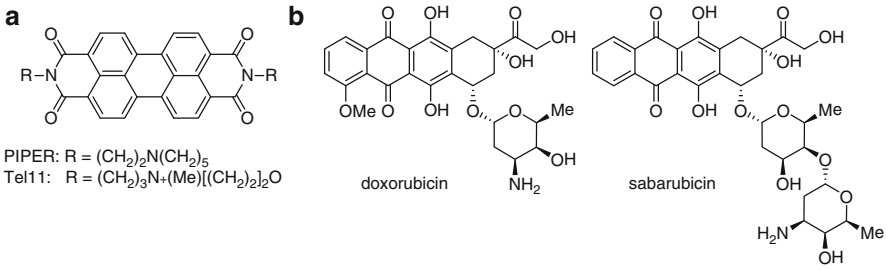


Fig. 52 (**a**) Representative perylene diimide derivatives. (**b**) Quadruplex-binding anthracycline antibiotics

3.3 *Perylene Diimide Derivatives*

Water-soluble perylene diimide derivatives, such as PIPER and analogues (Fig. 52a), represent a widely studied class of quadruplex-DNA ligands. Since the fluorescence of PIPER is quenched in aqueous solutions at neutral pH even in the absence of DNA due to strong self-aggregation, it is not particularly useful as a fluorescent probe [186]. However, an alternative study demonstrated significant quenching of fluorescence of PIPER (λ_{max} = 545–500 nm) in very diluted solutions (33.8 nM) upon binding to quadruplex-forming DNA sequences NHE-27[36] or 2HTR,[37] and the values of binding constants of 1.8×10^8 and 2.9×10^8 M^{-1} upon binding of seven or six molecules of ligand to these structures were obtained from the fluorimetric titrations [187].[38] Less aggregating analogues such as Tel11 are characterized by an intense fluorescence in more concentrated aqueous solutions (up to 1 μM), which is efficiently quenched upon interaction with quadruplex DNA. A nearly-complete quenching of Tel11 was obtained upon addition of substoichiometric amounts of 22AG, which indicates a positive cooperativity of binding, with apparent saturation upon binding of six molecules of the ligand per quadruplex [188]. However, it was shown that PIPER and Tel11 have rather similar affinity for duplex and quadruplex DNA [187, 189].

3.4 *Anthracyclines*

The anticancer antibiotics of the anthracycline series, doxorubicin and sabarubicin (Fig. 52b), are characterized by a fluorescence emission with maxima at 590 nm ($\Phi_F = 0.039$) and 565 nm ($\Phi_F = 0.063$), respectively. Their interaction with

[36] 5′-$TG_3GA\ G_3\ TG_4A\ G_3\ TG_4A_2\ G_2$-3′ (c-MYC oncogene).

[37] 5′-$A_2TC_2GTCGAGCAGAGT_2(AG_3T_2)_2$ AG-3′ (human telomeric repeat).

[38] In 10 mM Tris, 100 mM KCl, 1 mM EDTA.

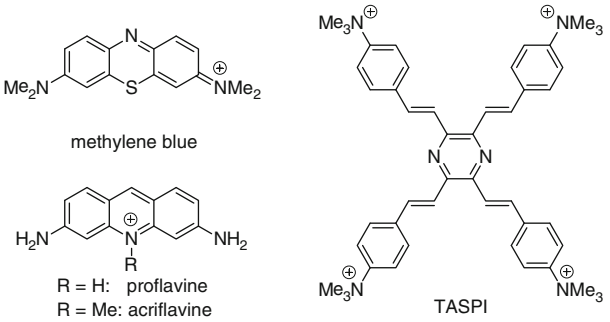


Fig. 53 Fluorescent dyes binding to quadruplex DNA with a decrease of fluorescence intensity

quadruplex DNA (human telomeric: d[$G_3(T_2AG_3)_3$]) results in efficient quenching of the fluorescence. In both cases, 1:1 drug–G4-DNA complexes are not fluorescent; instead the 2:1 complexes are fluorescent in Na^+-rich conditions ($\Phi_F = 0.009$ and 0.017 for doxorubicin and sabarubicin, respectively) but not in the K^+-rich buffer. The quadruplex-binding constants, calculated from the fluorimetric data, are log $K_{11} = 6.0$, log $K_{21} = 10.9$ for doxorubicin, and log $K_{11} = 5.8$, log $K_{21} = 11.5$ for sabarubicin in K^+ conditions[39]; similar values were obtained in Na^+-rich conditions. Notably, these values are slightly lower than those measured with duplex DNA (log K of 6.5 for doxorubicin and 6.1 for sabarubicin, respectively), thereby indicating the absence of quadruplex-vs-duplex selectivity of these drugs [190, 191].

3.5 *Classical Organic Dyes*

Several studies focused on the changes of fluorescent properties of well-established organic dyes, commonly used as stains in biochemistry and histology, upon their binding to quadruplex DNA. Thus, it was shown that methylene blue binds to intermolecular G-quadruplex DNA $(d[TTAAGG])_4$ with formation of 1:1 and 1:2 complexes (i.e., one cation of methylene blue stacked between two quadruplexes) with binding constants of 1.0×10^5 and 8.8×10^4 M^{-1}, respectively,[40] which leads to quenching of its fluorescence [192]. However, interaction with single-stranded and double-stranded DNA structures also leads to fluorescence quenching of methylene blue [193]. Binding of acridine dyes, such as proflavine and acriflavine, to cyclic diguanylic acid (c-di-GMP) was shown to induce formation of G-quadruplex structure in the absence of monovalent metal cations, which was accompanied by fluorescence quenching of these dyes (e.g., at $\lambda_{em} \approx 510$ nm for

[39] In 10 mM Tris, 1 mM EDTA, 50 mM KCl, pH 7.4.

[40] In 10 mM K_2HPO_4/KH_2PO_4, 1 mM EDTA, pH 7.04.

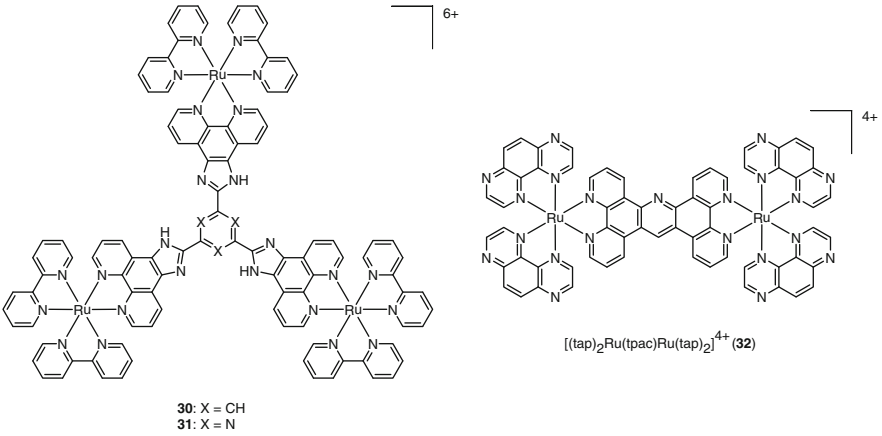


Fig. 54 Metal complexes displaying luminescence quenching upon binding to quadruplex DNA

proflavine) [194]. Interestingly, mononucleotides such as cGMP or rGTP did not induce fluorescence quenching of these acridine dyes; nonetheless, fluorescence quenching of proflavine upon binding to guanine-containing double-stranded DNA has been documented decades ago [195].

A four-branched pyrazine-styryl derivative TASPI (Fig. 53) was initially developed as a two-photon-absorbing dye, strongly fluorescent in aqueous solutions. However, the fluorescence is almost completely quenched upon addition of a stoichiometric amount of the thrombin-binding aptamer (TBA) quadruplex. Interestingly, addition of thrombin to the TASPI-TBA complex resulted in restoration of the fluorescence due to displacement of TASPI. This behavior has been applied to the selective fluorimetric detection of thrombin in the sub-micromolar concentration range [196].

3.6 Metal Complexes

In addition to the larger number of Ru(II) complexes whose luminescence increases upon interaction with quadruplex DNA (cf. Sect. 2.7), several complexes which act instead as "light-off" probes have been described. Like in the case of organic fluorophores, the quenching of the luminescent excited state of Ru(II) complexes is due to the photoinduced electron transfer from the guanine bases of quadruplex DNA to the ^{3}MLCT of the complex. Nice examples are provided by the trinuclear Ru(II) polypyridyl complexes **30** and **31** (Fig. 54), whose luminescence ($\lambda_{max} \approx 600$ nm) decreases by ~40% upon binding to 22AG (1:1 binding mode) [197], as well as the dinuclear $[(tap)_2Ru(tpac)Ru(tap)_2]^{4+}$ (**32**). In the latter case it was demonstrated that quadruplex DNA ($d(T_2AG_3)_4$) induces luminescence quenching much more efficiently than duplex DNA. Interestingly, this quenching behavior is in sharp contrast with the luminescence "light-up" effect of $[(phen)_2Ru(tppz)Ru$

$(phen)_2$], differing from **32** only by nitrogen substituents in the ligands (cf. Sect. 2.7.1). It was also shown that **32** is able to produce cross-links of guanine residues of the quadruplex upon photoillumination [198] (Fig. 54).

4 Permanent ("Tagged") Quadruplex-DNA Probes

Most of the above-mentioned studies rely on the intrinsic fluorescence of quadruplex binders or are based on rational design for adapting already known fluorophores to quadruplex recognition. Another approach consists of the covalent linking of a fluorescent moiety to a quadruplex-recognizing motif, resulting in a "tagged" quadruplex binder. This type of hybrid compound can also be called permanent probes as the fluorescence is generally not affected by the binding event. The main issue of this approach is the choice of the fluorescent tag that should not affect the binding features (affinity and selectivity) of the quadruplex-interactive moiety. In that sense, small-sized, neutral or poorly charged tags are recommended. Amongst the panel of dyes that fulfil these criteria and furthermore are chemically accessible and easily derivatizable we can find the boron-dipyrromethene (BODIPY) dye family.

Recently Nagasawa et al. developed a fluorescent derivative of telomestatin, the latter representing for many years the gold standard for quadruplex recognition. The macrocycle 7OTD (Fig. 55) is a synthetic analogue of telomestatin, which is more easily prepared and as efficient in terms of affinity and selectivity for quadruplexes in particular telomeric sequences. 7OTD was therefore a good candidate for functionalization by a fluorescent tag. L1BOD-7OTD exhibits a green fluorescence ($\lambda_{em} = 501$ nm) remaining unchanged in the absence or in the presence of various G4-forming sequences [199]. This stable fluorescence allowed the use of the labeled macrocycle to stain specifically G4-forming sequences in gel electrophoresis experiments (Fig. 56). Selective green staining of the telomeric quadruplex d$[T_2AG_3]_4$ was observed, whilst no fluorescence was detected with non-G4-forming sequences (duplex DNA or single-stranded sequence). Selective staining was also observed with other G4 forming sequences (c-myc, c-kit22, bcl2) and significant retardation of the band corresponding to the ligand–G4 complex is seen with increasing concentration of L1BOD-7OTD. Apparent dissociation constants of L1BOD-7OTD with various G4 forming sequences have been estimated by fluorescence polarization titration and shown to lie in the nanomolar range (0.2–80 nM depending on the G4 sequence). Finally, fixed HeLa cells were stained with L1BOD-7OTD that colocalizes with DAPI in the nuclei, indicating that the labeled compound can be used for imaging in cells.

Recently BODIPY has been used to label the bisquinolinum pyridodicarboxamide motif (PDC) affording the PDC-BODIPY compound (Fig. 57). In contrast to the PDC-conjugate built with TO as fluorescent moiety (PDC-TO see Sect. 2.1) the selectivity of PDC-BODIPY is retained as demonstrated by means of

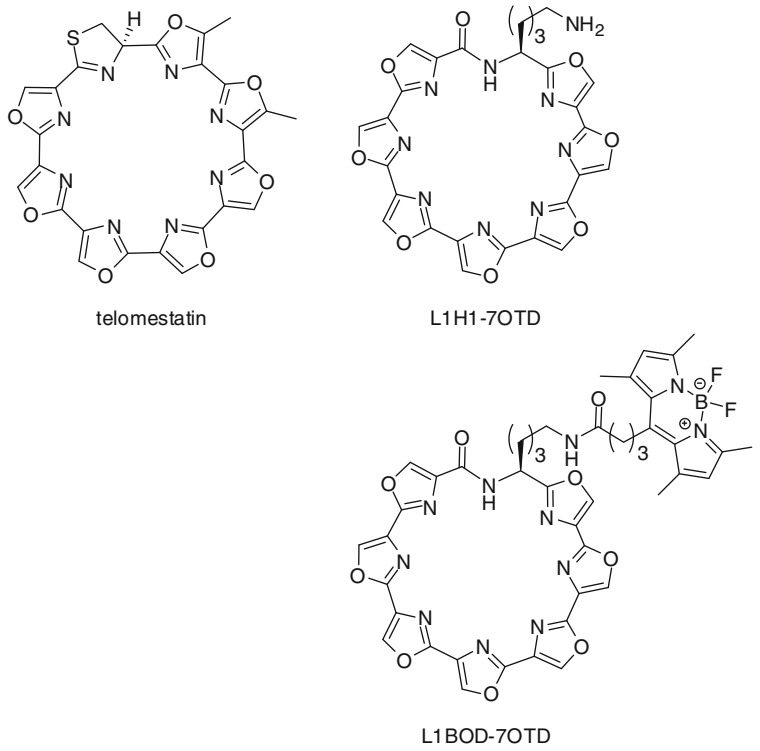


Fig. 55 Structures of telomestatin, L1H1-7OTD and the fluorescent labeled L1BOD-7OTD (LB)

a streptavidin-paramagnetic-particle based assay recently set up to evaluate optical selectivity of fluorescent G4-binders [200].

Finally, during the period of preparation of this manuscript, Rodriguez et al published an "in-situ" labeling strategy using click chemistry [201]. This result will not be described in detail but is nonetheless important with respect to this review. Click chemistry (i.e., the biocompatible 1,3-dipolar cycloaddition between azide-derivatized and alkyne-derivatized partners) has been successfully applied for fluorescent detection of replication and transcription in living cells using alkynyl uracil nucleotides and an azido derivative of Alexa Fluor 549 [202, 203]. The authors prepared an alkynyl derivative of pyridostatin [204]. The cells were first incubated with this compound and subsequently with an azido derivatized fluorophore (Alexa Fluor 549). The red fluorescence characteristic of Alexa Fluor 549 was found localized in the nucleoli with small foci in the nucleus. It was concluded that the compound accumulation in the nucleolar subcompartment results from preferential binding of alternative DNA and RNA structures. The latter may correspond to quadruplexes in consistency with the high GC rich content of ribosomal DNA and its transcriptional status in the nucleoli [205]. However, the rules that govern nucleolar targeting by small molecules are not well-established.

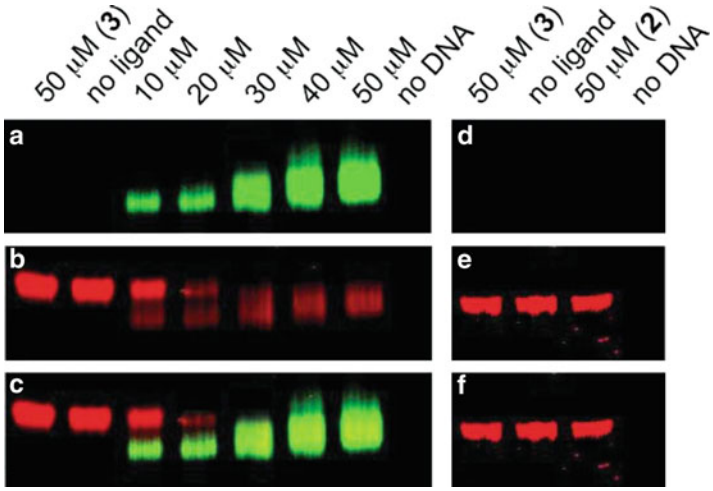


Fig. 56 Visualization of the telo24 G4 by L1BOD-7OTD by gel electrophoresis (12% polyacrylamide, 1X TBE buffer, 4 °C) in the presence of 10 μM oligonucleotides (**a–c**: telo 24, **d–f**: ds-telo 24) and various concentrations of L1BOD-7OTD. (*a*, *d*) All bands were detected using the 526 nm short pass filter. The gel was stained with (*b*) Stains-All and (*e*) ethidium bromide then all bands were detected using the 580–640 nm band pass filter. (*c*, *f*) Merged images of (*a*) and (*b*) or (*d*) and (*e*). Reprinted with permission from [199] with permission of The Royal Society of Chemistry

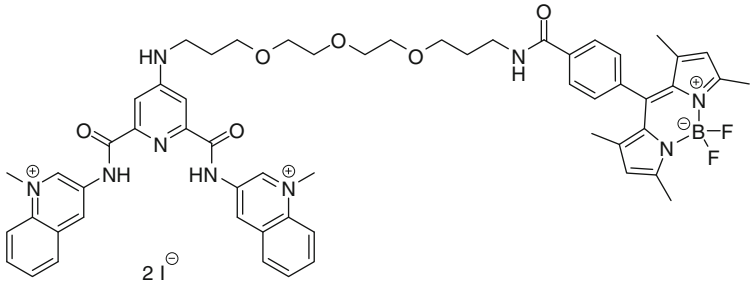


Fig. 57 Structure of PDC-BODIPY derivative

For instance, several DNA dyes (propidium iodide, thiazole orange) that intercalate in ds DNA are also known to label RNAs and to stain nucleoli; as well certain anticancer drugs (doxorubicin) with affinity for ds DNA have been shown to act in the nucleoli at the level of rRNA transcription. Thus the real significance of the nucleolar accumulation of G4 binders remains an open question. Finally, technical difficulties are not particularly addressed in the report, although "in-cell" click chemistry is notoriously known to require enormous amounts of copper salt ($CuSO_4$, 1 mM) that is strongly deleterious to cells. Nonetheless, from a more general point of view, this "in situ" labeling strategy opens highly interesting

perspectives and may allow one to circumvent the issues frequently associated with the labeling of active principles (i.e., modification of target binding and off target effects).

5 Conclusions and Perspectives

The selected examples above illustrate the tremendous growth of quadruplex-selective fluorescent ligands and new luminescence-based strategies in recent years. Considering the amount of literature devoted to the design and properties of quadruplex fluorescent probes one can easily appreciate the promising future for this class of agents.

A large number of quadruplex-selective probes are now available and many molecular designs have been explored. The choice of a particular probe design will be guided by the specific application of the study and instrumentation thereof, which in turn will determine the stringency of the chemical, photophysical and biocompatibility criteria required.

For instance, light-off probes are a priori less suitable for imaging in cells, whereas they may be of great interest for monitoring interactions and evaluating binding strength and selectivity for quadruplexes; they may also find application for various cation sensing in solution. In the case of light-up probes, the degree of signal enhancement for G4 vs duplex should be sufficient to allow this approach to be used in various situations. We have seen that this criterium is not always satisfied, although it is somewhat difficult to evaluate as it is often not clearly addressed or even obscured.

Although some dyes described in this review are only of academic interest, others seem to hold real promise in terms of practical applications. For instance, several scaffolds with exquisite quadruplex selectivity have been clearly identified, being already usable in various in vitro labeling experiments. Although in certain cases improvements are still needed in terms of solubility and self-aggregating aspects, we are confident that optimized second generations will appear in the near future.

We have seen that a number of probes active in vitro do not penetrate in living cells and thus have not been studied further although application in fixed cells could have been envisaged. A limited number have been used for imaging in cells or at the level of chromosomes. For those that are internalized in cells, most have been found in the nucleus and some more specifically sequestered in nucleoli. For the probes having a low in vitro affinity for duplex DNA, this might be associated with binding to quadruplex DNA or RNA structures. However it should be overstressed that whether the bound structures correspond to quadruplexes or to the many existing complex 3D RNA topologies remains to be determined.

Finally, particular attention should be paid to probes displaying a quadruplex-specific variation of their photophysical properties such as variable emission wavelength or distinct fluorescence lifetime. These agents can be useful tools for mapping quadruplex-containing domains, especially in view of the rapid

developments of ultrasensitive imaging instrumentation. In this regard it is worth mentioning that the performances of probes in term of brightness, photostability, and lifetime, represent currently the limit of super resolution and single molecule imaging, thereby opening a bright future for fluorescence engineering of quadruplex-responsive small molecules.

We have deliberately focused on reports addressing the potentiality of quadruplex-interactive small molecules for biological applications and have limited those concerning technological and nanomaterials applications. In addition, although fluorescently labeled oligonucleotides represent major players in the field and are widely used tools in cell biology, they are outside the scope of this review. Finally, every attempt has been made to ensure the comprehensiveness of this review but, due to the overwhelming number of publications in the field, omissions may have been made. The authors would like to apologize in advance for any incompleteness which might, therefore, result.

To conclude, the proposals mentioned here by no means represent the limits of the possible, but merely the most obvious next directions for the future development of the field.

Acknowledgments The authors acknowledge the *Centre National de la Recherche Scientifique* (CNRS), the Institut Curie, and the Fondation Pierre Gilles de Gennes for PhD and post-doctoral fellowships (E.L. and D.V.) We also thank the *Agence Nationale de la Recherche* (ANR-09-BLAN-0355 "G4 Toolbox") for funding.

References

1. Bacolla A, Wells RD (2009) Mol Carcinog 48:273
2. Cahoon LA, Seifert HS (2009) Science 325:764
3. Maizels N (2006) Nat Struct Mol Biol 13:1055
4. Paeschke K, Capra JA, Zakian VA (2011) Cell 145:678
5. Lopes J, Piazza A, Bermejo R, Kriegsman B, Colosio A, Teulade-Fichou M-P, Foiani M, Nicolas A (2011) EMBO J 30:4033
6. Broxson C, Beckett J, Tornaletti S (2011) Biochemistry 50:4162
7. Beaudoin J-D, Perreault J-P (2010) Nucleic Acids Res 38:7022
8. Davis JT (2004) Angew Chem Int Ed 43:668
9. Collie GW, Parkinson GN (2011) Chem Soc Rev 40:5867
10. Phan AT (2010) FEBS J 277:1107
11. Neidle S, Balasubramanian S (2006) RSC biomolecular sciences. The Royal Society of Chemistry, Cambridge, UK, p 316
12. Sun D, Thompson B, Cathers BE, Salazar M, Kerwin SM, Trent JO, Jenkins TC, Neidle S, Hurley LH (1997) J Med Chem 40:2113
13. Monchaud D, Teulade-Fichou M-P (2008) Org Biomol Chem 6:627
14. Georgiades SN, Abd Karim NH, Suntharalingam K, Vilar R (2010) Angew Chem Int Ed 49:4020
15. Balasubramanian S, Neidle S (2009) Curr Opin Chem Biol 13:345
16. Lipps HJ, Rhodes D (2009) Trends Cell Biol 19:414
17. Duquette ML, Handa P, Vincent JA, Taylor AF, Maizels N (2004) Genes Dev 18:1618
18. Neaves KJ, Huppert JL, Henderson RM, Edwardson JM (2009) Nucleic Acids Res 37:6269

19. Mela I, Kranaster R, Henderson RM, Balasubramanian S, Edwardson JM (2011) Biochemistry 51:578
20. Xu Y, Ishizuka T, Kurabayashi K, Komiyama M (2009) Angew Chem Int Ed 48:7833
21. Schaffitzel C, Berger I, Postberg J, Hanes J, Lipps HJ, Plückthun A (2001) Proc Natl Acad Sci USA 98:8572
22. Fernandez-Suarez M, Ting AY (2008) Nat Rev Mol Cell Biol 9:929
23. Kobayashi H, Ogawa M, Alford R, Choyke PL, Urano Y (2009) Chem Rev 110:2620
24. Neidle S (2009) Curr Opin Struct Biol 19:239
25. Patel DJ, Phan AT, Kuryavyi V (2007) Nucleic Acids Res 35:7429
26. Yang D, Okamoto K (2010) Future Med Chem 2:619
27. Collie GW, Sparapani S, Parkinson GN, Neidle S (2011) J Am Chem Soc 133:2721
28. Campbell NH, Parkinson GN, Reszka AP, Neidle S (2008) J Am Chem Soc 130:6722
29. Cosconati S, Marinelli L, Trotta R, Virno A, Mayol L, Novellino E, Olson AJ, Randazzo A (2009) J Am Chem Soc 131:16336
30. White EW, Tanious F, Ismail MA, Reszka AP, Neidle S, Boykin DW, Wilson WD (2007) Biophys Chem 126:140
31. Hamon F, Largy E, Guédin A, Rouchon-Dagois M, Sidibe A, Monchaud D, Mergny J-L, Riou J-F, Nguyen C-H, Teulade-Fichou M-P (2011) Angew Chem Int Ed 50:8745
32. Lusvarghi S, Murphy CT, Roy S, Tanious FA, Sacui I, Wilson WD, Ly DH, Armitage BA (2009) J Am Chem Soc 131:18415
33. Armitage BA (2005) In: Waring MJ, Chaires JB (eds) DNA binders and related subjects. Topics in Current Chemistry, vol 253. Springer, Berlin/Heidelberg, p 55
34. Sovenyhazy KM, Bordelon JA, Petty JT (2003) Nucleic Acids Res 31:2561
35. Nygren J, Svanvik N, Kubista M (1998) Biopolymers 46:39
36. Karunakaran V, Pérez Lustres JL, Zhao L, Ernsting NP, Seitz O (2006) J Am Chem Soc 128:2954
37. Lee LG, Chen C-H, Chiu LA (1986) Cytometry 7:508
38. Rye HS, Quesada MA, Peck K, Mathies RA, GIazer AN (1991) Nucleic Acids Res 19:327
39. Zhu H, Clark SM, Benson SC, Rye HS, Glazer AN, Mathies RA (1994) Anal Chem 66:1941
40. Netzel TL, Nafisi K, Zhao M, Lenhard JR, Johnson I (1995) J Phys Chem 99:17936
41. Monchaud D, Allain C, Teulade-Fichou M-P (2006) Bioorg Med Chem Lett 16:4842
42. Boger DL, Fink BE, Brunette SR, Tse WC, Hedrick MP (2001) J Am Chem Soc 123:5878
43. Boger DL, Tse WC (2001) Bioorg Med Chem 9:2511
44. Largy E, Hamon F, Teulade-Fichou M-P (2011) Anal Bioanal Chem 400:3419
45. Monchaud D, Allain C, Teulade-Fichou MP (2007) Nucleosides Nucleotides Nucleic Acids 26:1585
46. Monchaud D, Allain C, Bertrand H, Smargiasso N, Rosu F, Gabelica V, De Cian A, Mergny JL, Teulade-Fichou MP (2008) Biochimie 90:1207
47. Halder K, Largy E, Benzler M, Teulade-Fichou M-P, Hartig JS (2011) ChemBioChem 12:1663
48. Largy E, Saettel N, Hamon F, Dubruille S, Teulade-Fichou M-P (2012) Curr Pharm Des 18:1992
49. Lubitz I, Zikich D, Kotlyar A (2010) Biochemistry 49:3567
50. Allain C, Monchaud D, Teulade-Fichou M-P (2006) J Am Chem Soc 128:11890
51. Nakayama S, Kelsey I, Wang J, Roelofs K, Stefane B, Luo Y, Lee VT, Sintim HO (2011) J Am Chem Soc 133:4856
52. Haugland RP, The Molecular Probes Handbook: A Guide to Fluorescent Probes and Labeling Technologies (2002), 9th edn. Molecular Probes, Eugene, OR (USA)
53. Yang P, De Cian A, Teulade-Fichou MP, Mergny JL, Monchaud D (2009) Angew Chem Int Ed 48:2188
54. Lemarteleur T, Gomez D, Paterski R, Mandine E, Mailliet P, Riou JF (2004) Biochem Biophys Res Commun 323:802

55. Granotier C, Pennarun G, Riou L, Hoffschir F, Gauthier LR, De Cian A, Gomez D, Mandine E, Riou J-F, Mergny J-L, Mailliet P, Dutrillaux B, Boussin FD (2005) Nucleic Acids Res 33:4182
56. Lu Y-J, Ou T-M, Tan J-H, Hou J-Q, Shao W-Y, Peng D, Sun N, Wang X-D, Wu W-B, Bu X-Z, Huang Z-S, Ma D-L, Wong K-Y, Gu L-Q (2008) J Med Chem 51:6381
57. Lu Y-J, Yan S-C, Chan F-Y, Zou L, Chung W-H, Wong W-L, Qiu B, Sun N, Chan P-H, Huang Z-S, Gu L-Q, Wong K-Y (2011) Chem Commun 47:4971
58. Zipper H, Brunner H, Bernhagen J, Vitzthum F (2004) Nucleic Acids Res 32:e103
59. Xu H, Gao S, Yang Q, Pan D, Wang L, Fan C (2010) ACS Appl Mater Interfaces 2:3211
60. Xu Q-H, Wang S, Korystov D, Mikhailovsky A, Bazan GC, Moses D, Heeger AJ (2005) Proc Natl Acad Sci USA 102:530
61. Haugland RP (1996) Handbook of fluorescent probes and research chemicals. Molecular Probes, Eugene, OR
62. Kovalska V, Losytskyy M, Yarmoluk S, Lubitz I, Kotlyar A (2011) J Fluoresc 21:223
63. Yarmoluk S, Lukashova SS, Losytskyya MY, Akermanb B, Kornyushynac OS (2002) Spectrochim Acta A 58:3223
64. Ohulchanskyy TY, Pudavar HE, Yarmoluk SM, Yashchuk VM, Bergey EJ, Prasad PN (2007) Photochem Photobiol 77:138
65. Chen Q, Kuntz ID, Shafer RH (1996) Proc Natl Acad Sci USA 93:2635
66. Paramasivan S, Bolton PH (2008) Nucleic Acids Res 36:e106
67. Yang Q, Xiang J, Yang S, Zhou Q, Li Q, Tang Y, Xu G (2009) Chem Commun 2009:1103
68. Yang Q, Xiang J, Yang S, Li Q, Zhou Q, Guan A, Zhang X, Zhang H, Tang Y, Xu G (2010) Nucleic Acids Res 38:1022
69. Yang Q, Xiang J-F, Yang S, Li Q, Zhou Q, Guan A, Li L, Zhang Y, Zhang X, Zhang H, Tang Y, Xu G (2010) Anal Chem 82:9135
70. Meguellati K, Koripelly G, Ladame S (2010) Angew Chem Int Ed 49:2738
71. Koripelly G, Meguellati K, Ladame S (2010) Bioconjug Chem 21:2103
72. Latt SA, Stetten G (1976) J Histochem Cytochem 24:24
73. Demchenko AP (2009) In: Demchenko AP (ed) Introduction to fluorescence sensing. Springer, Netherlands, p 407
74. Maiti S, Chaudhury NK, Chowdhury S (2003) Biochem Biophys Res Commun 310:505
75. Phan AT, Kuryavyi V, Gaw HY, Patel DJ (2005) Nat Chem Biol 1:167
76. Jain AK, Reddy VV, Paul A, Muniyappa K, Bhattacharya S (2009) Biochemistry 48:10693
77. Jain AK, Bhattacharya S (2011) Bioconjug Chem 22:2355
78. Lepecq JB, Paoletti C (1967) J Mol Biol 27:87
79. Olmsted J, Kearns DR (1977) Biochemistry 16:3647
80. Ren J, Chaires JB (1999) Biochemistry 38:16067
81. Guo Q, Lu M, Marky LA, Kallenbach NR (1992) Biochemistry 31:2451
82. Koeppel F, Riou J-F, Laoui A, Mailliet P, Arimondo PB, Labit D, Petitgenet O, Hélène C, Mergny J-L (2001) Nucleic Acids Res 29:1087
83. Sun X, Cao E, He Y, Qin J (1999) Sci China Ser B Chem 42:62
84. Rosu F, De Pauw E, Guittat L, Alberti P, Lacroix L, Mailliet P, Riou J-F, Mergny J-L (2003) Biochemistry 42:10361
85. Martin MM, Plaza P, Meyer YH (1991) Chem Phys 153:297
86. Jurczok M, Plaza P, Martin MM, Rettig W (1999) J Phys Chem A 103:3372
87. Babendure JR, Adams SR, Tsien RY (2003) J Am Chem Soc 125:14716
88. Bhasikuttan AC, Mohanty J, Pal H (2007) Angew Chem Int Ed 46:9305
89. Mergny JL, Li J, Lacroix L, Amrane S, Chaires JB (2005) Nucleic Acids Res 33:e138
90. Kong DM, Ma YE, Wu J, Shen HX (2009) Chem Eur J 15:901
91. Guo J-H, Zhu L-N, Kong D-M, Shen H-X (2009) Talanta 80:607
92. Kong D-M, Ma Y-E, Guo J-H, Yang W, Shen H-X (2009) Anal Chem 81:2678
93. Kong D-M, Guo J-H, Yang W, Ma Y-E, Shen H-X (2009) Biosens Bioelectron 25:88

94. Ma DL, Kwan MHT, Chan DSH, Lee P, Yang H, Ma VPY, Bai LP, Jiang ZH, Leung CH (2011) Analyst 136:2692
95. Leung CH, Chan DS, Man BY, Wang CJ, Lam W, Cheng YC, Fong WF, Hsiao WL, Ma DL (2011) Anal Chem 83:463
96. Baptista MS, Indig GL (1998) J Phys Chem B 102:4678
97. Chang C-C, Wu JY, Chang TC (2003) J Chin Chem Soc 50:185
98. Chang C-C, Wu J-Y, Chien C-W, Wu W-S, Liu H, Kang C-C, Yu L-J, Chang T-C (2003) Anal Chem 75:6177
99. Chang C-C, Chien C-W, Lin Y-H, Kang C-C, Chang T-C (2007) Nucleic Acids Res 35:2846
100. Chang C-C, Kuo IC, Ling IF, Chen C-T, Chen H-C, Lou P-J, Lin J-J, Chang T-C (2004) Anal Chem 76:4490
101. Chang C-C, Chu J-F, Kao F-J, Chiu Y-C, Lou P-J, Chen H-C, Chang T-C (2006) Anal Chem 78:2810
102. Zhang X-F, Zhang H-J, Xiang J-F, Li Q, Yang Q-f, Shang Q, Zhang Y-X, Tang Y-L (2010) J Mol Struct 982:133
103. Dumat B, Bordeau G, Faurel-Paul E, Mahuteau-Betzer F, Saettel N, Bombled M, Metgé G, Charra F, Fiorini-Debuisschert C, Teulade-Fichou MP (2011) Biochimie 93:1209
104. Han FX, Wheelhouse RT, Hurley LH (1999) J Am Chem Soc 121:3561
105. Wei C, Jia G, Yuan J, Feng Z, Li C (2006) Biochemistry 45:6681
106. Zhang H-J, Wang X-F, Wang P, Ai X-C, Zhang J-P (2008) Photochem Photobiol Sci 7:948
107. Anantha NV, Azam M, Sheardy RD (1998) Biochemistry 37:2709
108. Zhang HJ, Wang XF, Wang P, Pang SP, Ai XC, Zhang JP (2008) Sci China Ser B Chem 51:452
109. Wei C, Han G, Jia G, Zhou J, Li C (2008) Biophys Chem 137:19
110. Wei C, Wang L, Jia G, Zhou J, Han G, Li C (2009) Biophys Chem 143:79
111. Wei C, Jia G, Zhou J, Han G, Li C (2009) Phys Chem Chem Phys 11:4025
112. Zhao P, Xu L-C, Huang J-W, Fu B, Yu H-C, Ji L-N (2009) Dyes Pigm 83:81
113. Nicoludis JM, Barrett SP, Mergny J-L, Yatsunyk LA (2012) Nucleic Acids Res 10.1093/nar/gks152
114. Arthanari H, Basu S, Kawano TL, Bolton PH (1998) Nucleic Acids Res 26:3724
115. Li T, Wang E, Dong S (2010) Anal Chem 82:7576
116. Li T, Dong S, Wang E (2010) J Am Chem Soc 132:13156
117. Hong Y, Lam JWY, Tang BZ (2011) Chem Soc Rev 40:5361
118. Hong Y, Häuβler M, Lam JWY, Li Z, Sin KK, Dong Y, Tong H, Liu J, Qin A, Renneberg R, Tang BZ (2008) Chem Eur J 14:6428
119. Hong Y, Xiong H, Lam JW, Haussler M, Liu J, Yu Y, Zhong Y, Sung HH, Williams ID, Wong KS, Tang BZ (2010) Chem Eur J 16:1232
120. Luo J, Xie Z, Lam JWY, Cheng L, Chen H, Qiu C, Kwok HS, Zhan X, Liu Y, Zhu D, Tang BZ (2001) Chem Commun 1740
121. Wang M, Zhang D, Zhang G, Tang Y, Wang S, Zhu D (2008) Anal Chem 80:6443
122. Huang J, Wang M, Zhou Y, Weng X, Shuai L, Zhou X, Zhang D (2009) Bioorg Med Chem 17:7743
123. Sanders MM (1998) Biochem Pharmacol 56:1157
124. Diaz MS, Freile ML, Gutierrez MI (2009) Photochem Photobiol Sci 8:970
125. Arora A, Balasubramanian C, Kumar N, Agrawal S, Ojha RP, Maiti S (2008) FEBS J 275:3971
126. Bhadra K, Kumar GS (2011) Biochim Biophys Acta 1810:485
127. Liu Y, Li B, Cheng D, Duan X (2011) Microchem J 99:503
128. Zhang W-J, Ou T-M, Lu Y-J, Huang Y-Y, Wu W-B, Huang Z-S, Zhou J-L, Wong K-Y, Gu L-Q (2007) Bioorg Med Chem 15:5493
129. Ma Y, Ou TM, Tan JH, Hou JQ, Huang SL, Gu LQ, Huang ZS (2009) Bioorg Med Chem Lett 19:3414

130. Franceschin M, Rossetti L, D'Ambrosio A, Schirripa S, Bianco A, Ortaggi G, Savino M, Schultes C, Neidle S (2006) Bioorg Med Chem Lett 16:1707
131. Tan J-H, Ou T-M, Hou J-Q, Lu Y-J, Huang S-L, Luo H-B, Wu J-Y, Huang Z-S, Wong K-Y, Gu L-Q (2009) J Med Chem 52:2825
132. Sun H, Tang Y, Xiang J, Xu G, Zhang Y, Zhang H, Xu L (2006) Bioorg Med Chem Lett 16:3586
133. Sun H, Xiang J, Tang Y, Xu G (2007) Biochem Biophys Res Commun 352:942
134. Juskowiak B, Galezowska E, Koczorowska N, Hermann TW (2004) Bioorg Med Chem Lett 14:3627
135. Alberti P, Schmitt P, Nguyen C-H, Rivalle C, Hoarau MS, Grierson D, Mergny J-L (2002) Bioorg Med Chem Lett 12:1071
136. Erkkila KE, Odom DT, Barton JK (1999) Chem Rev 99:2777
137. Keene FR, Smith JA, Collins JG (2009) Coord Chem Rev 253:2021
138. Kieltyka R, Englebienne P, Moitessier N, Sleiman H (2010) Methods Mol Biol (Clifton, NJ) 608:223
139. Jiang YL, Liu ZP (2010) Mini Rev Med Chem 10:726
140. Ma D-L, Che C-M, Yan S-C (2009) J Am Chem Soc 131:1835
141. Wu P, Ma D-L, Leung C-H, Yan S-C, Zhu N, Abagyan R, Che C-M (2009) Chem Eur J 15:13008
142. Yu C, Chan KH-Y, Wong KM-C, Yam VW-W (2009) Chem Commun 3756
143. Michael JC (2002) Coord Chem Rev 232:69
144. Jenkins Y, Friedman AE, Turro NJ, Barton JK (1992) Biochemistry 31:10809
145. Friedman AE, Chambron JC, Sauvage JP, Turro NJ, Barton JK (1990) J Am Chem Soc 112:4960
146. Shi S, Geng X, Zhao J, Yao T, Wang C, Yang D, Zheng L, Ji L (2010) Biochimie 92:370
147. Shi S, Zhao J, Geng X, Yao T, Huang H, Liu T, Zheng L, Li Z, Yang D, Ji L (2010) Dalton Trans 39:2490
148. Sun J, An Y, Zhang L, Chen HY, Han Y, Wang YJ, Mao ZW, Ji LN (2011) J Inorg Biochem 105:149
149. Rajput C, Rutkaite R, Swanson L, Haq I, Thomas JA (2006) Chem Eur J 12:4611
150. Pourtois G, Beljonne D, Moucheron C, Schumm S, Kirsch-De Mesmaeker A, Lazzaroni R, Brédas J-L (2003) J Am Chem Soc 126:683
151. Gill MR, Garcia-Lara J, Foster SJ, Smythe C, Battaglia G, Thomas JA (2009) Nat Chem 1:662
152. Shi S, Liu J, Yao T, Geng X, Jiang L, Yang Q, Cheng L, Ji L (2008) Inorg Chem 47:2910
153. Xu L, Zhang D, Huang J, Deng M, Zhang M, Zhou X (2010) Chem Commun 46:743
154. Goncalves DPN, Rodriguez R, Balasubramanian S, Sanders JKM (2006) Chem Commun 2006:4685
155. Ren L, Zhang A, Huang J, Wang P, Weng X, Zhang L, Liang F, Tan Z, Zhou X (2007) ChemBioChem 8:775
156. Alzeer J, Vummidi BR, Roth PJC, Luedtke NW (2009) Angew Chem Int Ed 48:9362
157. Qin H, Ren J, Wang J, Luedtke NW, Wang E (2010) Anal Chem 82:8356
158. Ren J, Qin H, Wang J, Luedtke N, Wang E, Wang J (2011) Anal Bioanal Chem 399:2763
159. Alzeer J, Luedtke NW (2010) Biochemistry 49:4339
160. Cha A, Snyder GE, Selvin PR, Bezanilla F (1999) Nature 402:809
161. Caravan P, Ellison JJ, McMurry TJ, Lauffer RB (1999) Chem Rev 99:2293
162. Bünzli JC (2010) Chem Rev 110:2729
163. Fu PKL, Turro C (1999) J Am Chem Soc 121:1
164. Zhang H, Yu H, Ren J, Qu X (2006) Biophys J 90:3203
165. Lis S, Elbanowski M, Mąkowska B, Hnatejko Z (2002) J Photochem Photobiol, A 150:233
166. Lehn J-M (1990) Angew Chem Int Ed 29:1304

167. Galezowska E, Gluszynska A, Juskowiak B (2007) J Inorg Biochem 101:678
168. Worlinsky JL, Basu S (2009) J Phys Chem B 113:865
169. Xu H, Zhang H, Qu X (2006) J Inorg Biochem 100:1646
170. Knobloch B, Linert W, Sigel H (2005) Proc Natl Acad Sci USA 102:7459
171. Lee J, Park M, Son HS, Lee SB, Lee HC, Ku JK, Park JW (2002) Biopolymers 67:413
172. Xie H, Yang D, Heller A, Gao Z (2007) Biophys J 92:L70
173. Delaney S, Barton JK (2003) Biochemistry 42:14159
174. Teulade-Fichou M-P, Carrasco C, Guittat L, Bailly C, Alberti P, Mergny J-L, David A, Lehn J-M, Wilson WD (2003) J Am Chem Soc 125:4732
175. Mergny JL, Lacroix L, Teulade-Fichou MP, Hounsou C, Guittat L, Hoarau M, Arimondo PB, Vigneron JP, Lehn JM, Riou JF, Garestier T, Helene C (2001) Proc Natl Acad Sci USA 98:3062
176. Peduto A, Pagano B, Petronzi C, Massa A, Esposito V, Virgilio A, Paduano F, Trapasso F, Fiorito F, Florio S, Giancola C, Galeone A, Filosa R (2011) Bioorg Med Chem 19:6419
177. Cummaro A, Fotticchia I, Franceschin M, Giancola C, Petraccone L (2011) Biochimie 93:1392
178. Franceschin M, Ginnari-Satriani L, Alvino A, Ortaggi G, Bianco A (2010) Eur J Org Chem 2010:134
179. Petraccone L, Fotticchia I, Cummaro A, Pagano B, Ginnari-Satriani L, Haider S, Randazzo A, Novellino E, Neidle S, Giancola C (2011) Biochimie 93:1318
180. Xu Y, Yamazaki S, Osuga H, Sugiyama H (2006) Nucleic Acids Symp Ser 50:183
181. Shinohara K-I, Sannohe Y, Kaieda S, Tanaka K-I, Osuga H, Tahara H, Xu Y, Kawase T, Bando T, Sugiyama H (2010) J Am Chem Soc 132:3778
182. Çetinkol ÖP, Engelhart AE, Nanjunda RK, Wilson WD, Hud NV (2008) ChemBioChem 9:1889
183. Xu W, Tan JH, Chen SB, Hou JQ, Li D, Huang ZS, Gu LQ (2011) Biochem Biophys Res Commun 406:454
184. Granzhan A, Ihmels H, Jager K (2009) Chem Commun 2009:1249
185. Cheng MK, Modi C, Cookson JC, Hutchinson I, Heald RA, McCarroll AJ, Missailidis S, Tanious F, Wilson WD, Mergny JL, Laughton CA, Stevens MF (2008) J Med Chem 51:963
186. Kerwin SM, Chen G, Kern JT, Thomas PW (2002) Bioorg Med Chem Lett 12:447
187. Samudrala R, Zhang X, Wadkins RM, Mattern DL (2007) Bioorg Med Chem 15:186
188. Kern JT, Kerwin SM (2002) Bioorg Med Chem Lett 12:3395
189. Mazzitelli C, Brodbelt J, Kern J, Rodriguez M, Kerwin S (2006) J Am Soc Mass Spectrom 17:593
190. Manet I, Manoli F, Zambelli B, Andreano G, Masi A, Cellai L, Monti S (2011) Phys Chem Chem Phys 13:540
191. Manet I, Manoli F, Zambelli B, Andreano G, Masi A, Cellai L, Ottani S, Marconi G, Monti S (2011) Photochem Photobiol Sci 10:1326
192. Sun H, Xiang J, Zhang Y, Xu G, Xu L, Tang Y (2006) Chin Sci Bull 51:1687
193. Tong C, Hu Z, Wu J (2010) J Fluoresc 20:261
194. Nakayama S, Kelsey I, Wang J, Sintim HO (2011) Chem Commun 47:4766
195. Thomes JC, Weill G, Daune M (1969) Biopolymers 8:647
196. Yan S, Huang R, Zhou Y, Zhang M, Deng M, Wang X, Weng X, Zhou X (2011) Chem Commun 47:1273
197. Xu L, Liao GL, Chen XA, Zhao CY, Chao H, Ji LN (2010) Inorg Chem Commun 13:1050
198. Rickling S, Ghisdavu L, Pierard F, Gerbaux P, Surin M, Murat P, Defrancq E, Moucheron C, Kirsch-De Mesmaeker A (2010) Chem Eur J 16:3951
199. Tera M, Iida K, Ikebukuro K, Seimiya H, Shin-Ya K, Nagasawa K (2010) Org Biomol Chem 8:2749
200. Largy E, Hamon F, Teulade-Fichou M-P Methods, in press (DOI: 10.1016/j.ymeth.2012.02.008)
201. Rodriguez R, Miller KM, Forment JV, Bradshaw CR, Nikan M, Britton S, Oelschlaegel T, Xhemalce B, Balasubramanian S, Jackson SP (2012) Nat Chem Biol 8:301

202. Jao CY, Salic A (2008) Proc Natl Acad Sci USA 105:15779
203. Salic A, Mitchison TJ (2008) Proc Natl Acad Sci USA 105:2415
204. Rodriguez R, Müller S, Yeoman JA, Trentesaux C, Riou J-F, Balasubramanian S (2008) J Am Chem Soc 130:15758
205. Drygin D, Siddiqui-Jain A, O'Brien S, Schwaebe M, Lin A, Bliesath J, Ho CB, Proffitt C, Trent K, Whitten JP, Lim JKC, Von Hoff D, Anderes K, Rice WG (2009) Cancer Res 69:7653
206. Wilson T, Williamson MP, Thomas JA (2010) Organic & Biomolecular Chemistry 8:2617

Top Curr Chem (2013) 330: 179–210
DOI: 10.1007/128_2012_351

Published online: 11 August 2012

Calculation of Hydrodynamic Properties for G-Quadruplex Nucleic Acid Structures from *in silico* Bead Models

Huy T. Le, Robert Buscaglia, William L. Dean, Jonathan B. Chaires, and John O. Trent

Abstract Nucleic acids enriched in guanine bases can adopt unique quadruple helical tertiary structures known as G-quadruplexes. G-quadruplexes have emerged as attractive drug targets as many G-quadruplex-forming sequences have been discovered in functionally critical sites within the human genome, including the telomere, oncogene promoters, and mRNA processing sites. A single G-quadruplex-forming sequence can adopt one of many folding topologies, often resulting in a lack of a single definitive atomic-level resolution structure for many of these sequences and a major challenge to the discovery of G-quadruplex-selective small molecule drugs. Low-resolution techniques employed to study G-quadruplex structures (e.g., CD spectroscopy) are often unable to discern between G-quadruplex structural ensembles, while high-resolution techniques (e.g., NMR spectroscopy) can be overwhelmed by a highly polymorphic system. Hydrodynamic bead modeling is an approach to studying G-quadruplex structures that could bridge the gap between low-resolution techniques and high-resolution molecular models. Here, we present a discussion of hydrodynamic bead modeling in the context of studying G-quadruplex structures, highlighting recent successes and limitations to this approach, as well as an example featuring a G-quadruplex structure formed from the human telomere. This example can easily be adapted to the investigation of any other G-quadruplex-forming sequences.

Keywords Bead models · Drug discovery · G-quadruplex · Hydrodynamic · Nucleic acids · Sedimentation

H.T. Le, R. Buscaglia, W.L. Dean, J.B. Chaires and J.O. Trent (✉)
Clinical and Translation Research Building, University of Louisville, 505 S. Hancock St, Louisville, KY 40202, USA
e-mail: john.trent@louisville.edu

Contents

1 Introduction 180
2 G-Quadruplex Nucleic Acids 180
2.1 G-Quadruplex Structures as Drug Targets 180
2.2 Elucidating the Structures of G-Quadruplex-Forming Sequences 182
3 Hydrodynamic Bead Modeling of G-Quadruplex Structures 186
3.1 History and Overview 186
3.2 Calibration of HYDROPRO Parameters 190
4 Limitations of Hydrodynamic Bead Modeling 195
4.1 Hydrodynamic Bead Modeling is a Low Resolution Technique 195
4.2 Treatment of Hydration in Hydrodynamic Bead Modeling 196
4.3 Potassium Binding by G-Quadruplex and Its Implication on Hydrodynamics 198
5 Future Directions for G-Quadruplex Hydrodynamic Bead Modeling 201
6 Conclusion 204
References 204

1 Introduction

Hydrodynamic bead modeling has emerged as a useful tool for structural biologists studying biological macromolecules and their complexes for which high-resolution structural coordinates are unavailable [1]. The comparison of directly measurable properties with theoretical calculations on structures and molecular models can provide fundamental insight, albeit at low resolution, into structures in solution. One particular macromolecular family that we and others have applied this technique to is G-quadruplex, a unique type of tertiary structure formed from guanine-rich DNA and RNA sequences. We will present a brief introduction to G-quadruplex nucleic acids focusing on some of the underlying challenges experienced when studying these structures at atomic-level resolution, followed by a discussion of hydrodynamic bead modeling in the study of G-quadruplex structures, and then conclude with some of the limitations and possible future directions of the application of this technique. Here, we focus on aspects of hydrodynamics as it relates to G-quadruplex structures. For a general overview of hydrodynamics history, theories, and modeling techniques, the readers will benefit from several excellent reviews available from Byron [1], Carrasco and García de la Torre [2], and García de la Torre and Bloomfield [3].

2 G-Quadruplex Nucleic Acids

2.1 G-Quadruplex Structures as Drug Targets

The field of drug discovery is largely protein-centric but there are efforts to find and target non-protein systems. A recent assessment of potential drug targets concluded that approximately only 10–15% of the human proteome is “druggable” [4], which

is defined as the intersection of sets of proteins that are capable of binding "drug-like" molecules and proteins that are products of disease modifying genes. The total number of potentially viable protein targets, thus, may be very small and so it is critical to consider options for drug discovery that involve other biomolecules. DNA and RNA are fundamentally attractive yet underrepresented and underutilized alternatives to proteins as drug targets [4–6]. In the progression through the central dogma, transcription of DNA produces RNA and translation of RNA produces proteins. At each stage of the process the absolute number of targets that may be hit by a drug molecule increases. Aiming at a particular drug target at the level of gene expression by targeting DNA and RNA rather than the numerous resultant proteins could allow for more efficient drug action. However, targeting genetic DNA sequence selectively via small molecules [7–9] is extremely challenging, as is using oligonucleotide-based approaches such as the antigene, antisense, and iRNA which have their own set of challenges [10–15].

DNA and RNA are polymorphic and can adopt a variety of other secondary and tertiary arrangements in addition to the standard Watson–Crick double helix. Often these unique structures are localized to functionally important sites known to play a role in regulating gene expression. One of these particular tertiary structures is the G-quadruplex, a unique quadruple helix formed from the self-assembly of guanine-rich DNA and RNA sequences containing four or more runs of at least two guanines. A G-quadruplex usually consists of two or three stacked G-quartets or tetrads, which involve four guanines in a square planar arrangement that are stabilized by Hoogsteen hydrogen bonds. The presence of a monovalent cation in the center of the G-quartet can greatly increase the stability of the structure (stability: $Li^+ < Na^+ < K^+$) [16]. G-quadruplexes can form from the unimolecular folding of a single-strand, or form by the bimolecular or tetramolecular association of strands.

The human genome is believed to contain more than 370,000 putative G-quadruplex forming sequences [17, 18]. These sequences have been discovered throughout the genome [17] and are found to be localized to functionally important sites. Amongst these sites is the human telomere [19], a 5,000–8,000 base pair long sequence of DNA consisting of the repeats $d(T_2AG_3)$ at the end of the human chromosome. The telomere, in addition to serving as the biological clock for the cell, also contributes to genetic stability by capping and protecting the chromosomes from damage during replication [20]. The formation of G-quadruplex is believed to occur at the extreme 3′ end of the telomere that contains a single-stranded overhang of 100–200 bases [21]. Small molecules that stabilize G-quadruplexes formed from telomeric oligonucleotide sequences *in vitro* have been shown to decrease the activity of telomerase, an enzyme which is responsible for maintaining the length of telomeric DNA, in addition to leading to a DNA double-stranded break response by the cell that results in cellular senescence and cell death [20, 22–25]. Since telomerase activation has been found to be involved in greater than 90% of all cancers [26], G-quadruplex formation in the human telomere is an attractive anti-cancer drug target. In addition to the human telomere, G-quadruplex-forming sequences is overrepresented in the promoter regions of many oncogenes, such as *c-Myc* [27],

c-Kit [28, 29], *Bcl-2* [30], *VEGF* [31], and *HIF-1α* [32], whereas the occurrence of G-quadruplex-forming sequences in tumor suppressor genes tends to be much lower [33]. These sequences were found to be evolutionary conserved between humans and related species [34]. The formation of G-quadruplex structures in these promoter regions is believed to influence gene transcription [35, 36]. In the case of the putative G-quadruplex-forming sequence located upstream of the P1 promoter for the *c-Myc* proto-oncogene, G-quadruplex formation was demonstrated to be the primary regulator for *c-Myc* expression [37]. Last, but not least, putative G-quadruplex-forming sequences have also been found in the 5′-untranslated or 3′-untranslated regions of about 20% of genes [38] and were discovered to regulate gene expression by repressing translation [39, 40] or promoting alternative splicing [41]. These findings seem to suggest that G-quadruplexes could play a key role in fundamental biological processes essential for normal cellular functions as well as cell growth and differentiation.

2.2 *Elucidating the Structures of G-Quadruplex-Forming Sequences*

Over four decades before the discovery of the double helix, high concentrations of guanine were observed to form gels in aqueous solution [42]. The structure involved was later determined in 1962, 9 years after Watson and Crick presented their findings on the double helix, to be the G-quartet or tetrad (Fig. 1a) [43]. The double helical structure discovered by Watson and Crick, known as B-DNA, is one of several forms that DNA can adopt. In addition to the canonical B-DNA, two other major forms, A-DNA and Z-DNA have been studied extensively. A recent review revealed that many other forms of DNA have been reported in the literature [44]. However, most of these DNA structures were observed *in vitro* and it is unclear what biological relevance, if any, they possess. While the structures of double helical nucleic acids appear to be highly polymorphic, this polymorphism pales in comparison to the possible structural diversity that can be assumed by guanine quadruple helices. G-quadruplexes can be classified primarily based on molecularity with a decrease in molecularity associated with an increase in the structural polymorphism (Fig. 1b–d). Tetramolecular G-quadruplexes, formed from four guanine-rich DNA or RNA sequences, are the least polymorphic, adopting only a parallel conformation with all four strands in the same direction (Fig. 1b) [45, 46]. Bimolecular G-quadruplexes, formed from the dimerization of two guanine-rich DNA or RNA sequences folded into double-stranded helical duplex structures, are more diverse with three possible conformations, one parallel and two antiparallels (Fig. 1c). In the parallel conformation the two strands of the duplex are connected by a double chain-reversal, or propeller, loop while in the antiparallel conformation the two strands are connected by a hairpin loop – a diagonal loop if the interconnecting strands oppose one another within the G-quadruplex structure

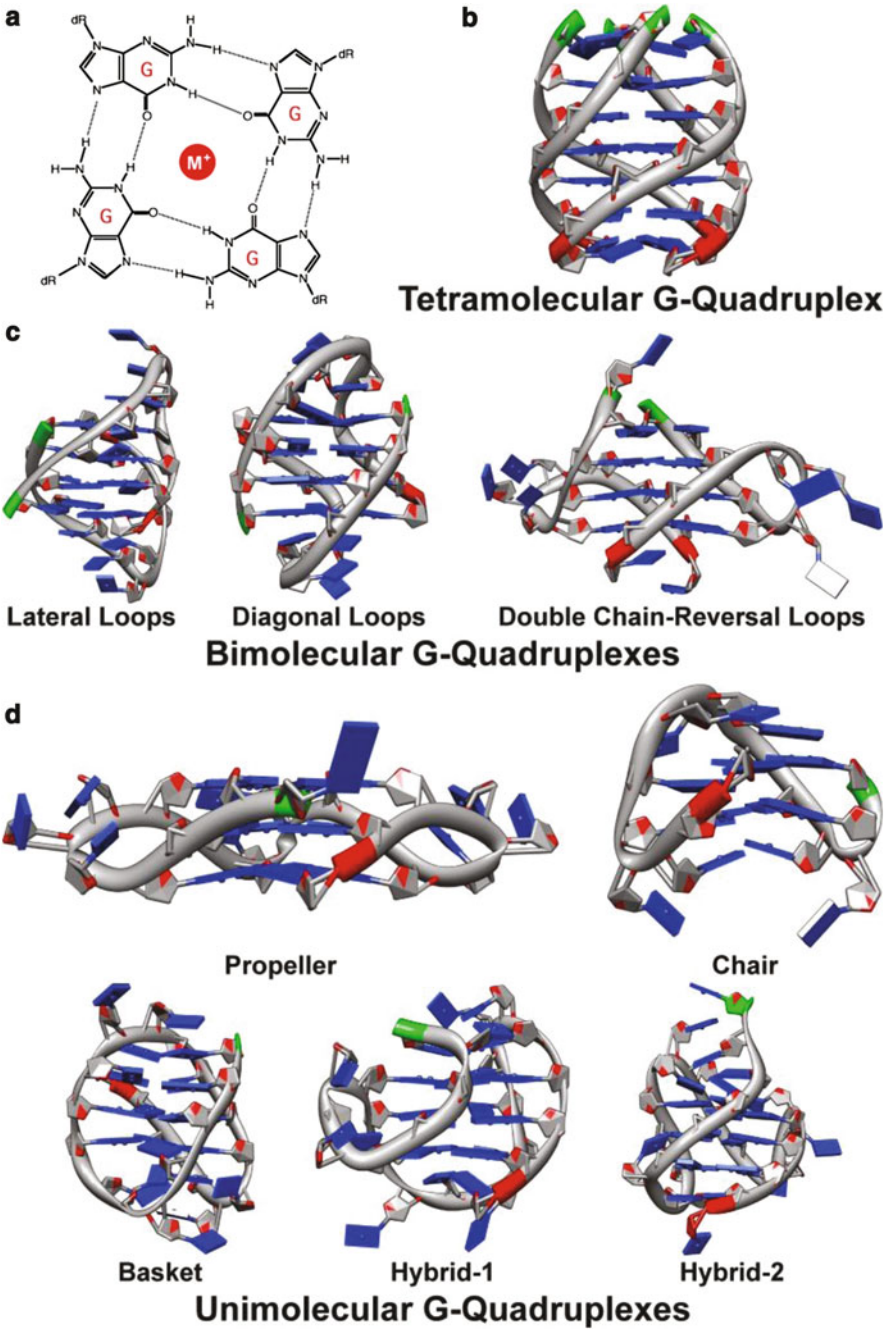


Fig. 1 G-quadruplex structures and polymorphism. The formation of a G-quartet (**a**), a square planar arrangement of four guanine bases stabilized by centrally coordinating monovalent cation with Hoogsteen hydrogen bonds between guanine bases, is the first step in the polymorphic formation of G-quadruplex structures (**b**, **c**, **d**). In (**b**), a tetramolecular G-quadruplex, from four

[47–50] or a lateral, or edgewise, loop if the strands are adjacent to one another [51]. Lastly, unimolecular G-quadruplexes, formed from the folding of a single guanine-rich DNA or RNA sequence into a four-stranded quadruple helix structure with three connecting loops, are the most polymorphic, capable of adopting multiple topologies (Fig. 1d) depending on the condition.

Much of our understanding of the polymorphic nature of G-quadruplex structures comes from structural studies involving G-quadruplex-forming sequences derived from the human telomere. In 1993, the first structure of a unimolecular G-quadruplex formed from a human genomic sequence was solved by Wang and Patel (PDB: 143D) [52]. It was discovered by NMR spectroscopy that the human telomeric sequence $dAG_3(T_2AG_3)_3$ in Na^+ solution folds into an anti-parallel G-quadruplex topology consisting of three stacked G-quartets with two lateral and a diagonal connecting loops. The structure in K^+ solution, however, proved to be more difficult. In 2002, a structure for the human telomeric sequence in K^+ was determined by X-ray crystallography (PDB: 1KF1) [53]. Parkinson *et al.* reported that in K^+ the human telomeric sequence adopted a parallel topology consisting of three stacked G-quartets and three double chain-reversal loops. However, the reported crystal structure did not agree with biophysical measurements, including sedimentation velocity and equilibrium experiments, CD spectroscopy measurements, and 2-aminopurine fluorescent quenching studies in solution, and in 2005 Li *et al.*, using a combined approach of hydrodynamic bead modeling and experimental measurement, demonstrated conclusively that the parallel structure was not the predominant form adopted by the human telomeric sequence in K^+ solution [54]. It was suggested in the same study that perhaps the crowding or dehydrated conditions in the crystals led to the parallel conformation. This hypothesis was confirmed when, in 2010, Miller *et al.*, using CD and NMR spectroscopy, determined that dehydration not molecular crowding determines the conformation of G-quadruplex structure [55]. Shortly after, in 2011, Heddi and Phan solved the NMR solution structure (PDB: 2LD8) for the same sequence in 40% polyethylene glycol (PEG 200) and discovered that the structure formed was close to the previously reported crystal structure [56]. After the 2005 Li *et al.* report that the solution structure of the human telomere was not predominantly the crystal form, alternative folding topologies were discovered. In addition to the antiparallel and the parallel conformations, two mixed topologies (i.e., three strands in one direction, one strand in the opposite direction) were reported for the human telomeric sequence in K^+ solution by NMR spectroscopy. The first conformation,

Fig. 1 (continued) guanine-rich DNA or RNA strands, assumes a parallel topology (PDB ID: 139D). In (**c**), bimolecular G-quadruplexes, from two guanine-rich sequences, can form either a parallel structure with two double chain reversal loops (3CE5) or one of two possible anti-parallel topologies with either two lateral loops (1D59) or two diagonal loops (1JRN). In (**d**), G-quadruplex structures are formed from the folding of a single guanine-rich sequence into a quadruple helical structure that can assume one of several topologies including but not limited to propeller (1KF1), chair (148D), basket (143D), hybrid-1 (2GKU), or hybrid-2 (2JPZ). In all structures, the 5′ end is colored *green* and the 3′ end is colored *red*

hybrid-1, was reported by Ambrus *et al.* in April 2006 but they did not publish the full structure [57]. Shortly after, Luu *et al.* in July 2006 reported the full NMR structure for the hybrid-1 topology consisting of a three stack G-quadruplex with a chain reversal loop followed by two lateral loops (PDB: 2GKU) [58]. The second mixed conformation, hybrid-2, reported by Dai *et al.* later in 2007, differed from the hybrid-1 conformation by the order of the connecting loops, two lateral loops followed by a chain reversal loop (PDB: 2JPZ) [59]. In addition, Lim *et al.* reported in 2009 an antiparallel structure of the human telomeric sequence in K^+ solution (PDB: 2KF8) that was very similar to the structure in Na^+ solution [60] but which contained only two stacked G-quartets in K^+ as opposed to three stacked G-quartets in Na^+ solution. To expand the observed folding topologies further, another antiparallel folding topology, this one consisting of three lateral loops, was also been reported [61]. However, this topology has only been seen with the thrombin binding aptamer and has not been observed for the human telomere or any other human genomic sequences.

As we and others have reported, the folding of a single DNA or RNA strand into a unimolecular G-quadruplex can be highly polymorphic [62, 63]. The human telomeric sequence, while it is a relatively simple G-quadruplex-forming sequence, can potentially fold into more than 200 intramolecular conformations [62] with variations in connecting loop types (i.e., lateral, diagonal, or double chain-reversal), strand/segment orientations (i.e., parallel, antiparallel, or hybrid 3+1), numbers of G-quartets (i.e., two or three), and glycosyl torsion angles (i.e., *syn* or *anti*) depending on the conditions (e.g., buffer composition, the presence of organic solvents such as acetonitrile or ethanol, DNA concentration, ion concentration, annealing profile, and the presence of various biological molecules). Consequently, when a G-quadruplex-forming sequence is studied, certain steps are usually taken to reduce this inherent polymorphism. The most common approach for reducing the structural polymorphism of G-quadruplex-forming sequences is sequence modification [64, 65]. A sequence is usually modified through a series of base subtraction, addition, and/or substitution to yield a new modified sequence with reduced polymorphism with the ultimate goal of sufficient enrichment of one species for NMR structure elucidation. Sequence modifications often include the incorporation non-canonical bases, such as 8-methylguanine or 8-bromoguanine, which are known to produce G-quadruplex structures with a *syn* glycosidic configuration [66–68], while use of O6-methylguanine, inosine, or 6-thioguanine has been shown to destabilize G-quadruplex formation [69–72]. Substitution of 8-aminoguanine promotes formation of tetramolecular parallel quadruplexes such as those formed by TG_4T [73]. Modifications of the sugar-phosphate backbone by insertion of 5′–5′ or 3′–3′ polarity inversion have also been shown to have a dramatic effect on G-quadruplex formation and stability [74] and use of RNA or LNA forces adoption of a *syn* glycosidic guanosine conformation [75–78]. In some cases, sequence modification can result in new sequences that differ only slightly from the parent sequence, e.g., the human telomere sequence [58–60, 79–81], while often the resulting sequences differ significantly from the parent sequence, e.g., the *c-Myc* promoter sequence [27, 82–85]. In addition to sequence modifications, another common approach to

reduce polymorphism is by changing the solution conditions. The addition of biological molecules [86] (e.g., sugar, proteins), presence of co-solvents (e.g., acetonitrile, PEG) [55, 87], choice of divalent vs monovalent cations [88, 89], and cation concentration [90, 91] all play a major role in directing G-quadruplex formation and determining stability. Due to the significant influence sequence and conditions have on G-quadruplex population distribution, it is very important that a complete description of the experimental methods/conditions used in sample preparation be fully reported for reproducibility.

One potential consequence in using any polymorphism selection method described above for altering the distribution of the polymorphic states of the G-quadruplex structures is that such an approach can result in an unpredictable perturbation of the system which may or may not be present in the original population of states (this has a significant effect on what can be claimed as "biologically relevant"). In the case of the human telomeric sequence, it was demonstrated that small changes in the flanking base of the G-quadruplex-forming segment $dG_3(T_2AG_3)_3$, (-core-), resulted in dramatically different topologies, such as hybrid-1 (2GKU – 5′-T_2-core-A-3′, 2HY9 – 5′-A_3-core-A_2-3′, and 2JSM – 5′-TA-core-3′) [58, 79, 80], hybrid-2 (2JPZ – 5′-T_2A-core-T_2-3′, 2JSL – 5′-TA-core-T_2-3′) [59, 80], and the antiparallel conformation (2KF8 – 5′-core-T-3′, and 2KKA – 5′-A-core-T-3′) [60, 81]. The untested assumption is that these means of reducing the polymorphic distribution select a member of the ensemble of species originally formed by the parent sequence. In fact, such perturbation can shift the equilibrium to favor species that might not be indicative of, or be completely dissimilar to, any species actually formed *in vitro* or *in vivo* [62]. Further, G-quadruplex polymorphism has severely hindered investigation of quadruplex structure, biophysical properties, small molecule interaction, and the thermodynamics of quadruplex formation. Techniques commonly used to study G-quadruplex DNA are typically too low in resolution to distinguish between species in a mixture, e.g., CD spectroscopy, UV–vis spectroscopy, ultracentrifugation, and gel electrophoresis, while high resolution techniques, e.g., NMR spectroscopy, are of limited usefulness for mixtures containing multiple G-quadruplex topologies [63].

3 Hydrodynamic Bead Modeling of G-Quadruplex Structures

3.1 History and Overview

Several approaches have been proposed for studying G-quadruplex structures without perturbing the equilibrium of the G-quadruplex species in solution by sequence modification. In one approach, G-quadruplex species in solution are

separated using size-exclusion chromatography and individual species are studied using a combination of biophysical techniques (e.g., analytical ultracentrifugation, CD spectroscopy, UV–vis spectroscopy, NMR spectroscopy) [63, 92]. This approach discovered that the *c-Kit* promoter G-quadruplex-forming sequence exists in solution as a mixture of monomers, dimers, and higher-order G-wire structures [92]. The disadvantage of this approach is that it relies on chromatographic separation and is of limited usefulness for some sequences, such as the human telomeric sequence, that elute and sediment essentially as one species. In such instances, another approach employing hydrodynamic bead modeling has emerged as a useful tool for studying G-quadruplex structures at atomic-level resolution. The formulation and development of hydrodynamic bead modeling came about from the extensive theoretical work of Kirkwood, Riseman, Bloomfield, García de la Torre, and Allison. Their research is discussed in several extensive reviews and the readers are referred to these articles for a more in depth discussion of the topic [1–3]. Briefly, the structure of a macromolecule is represented as a collection of solid spheres (i.e., beads on a string model), for which known hydrodynamic properties can be calculated. A sphere can represent an individual atom, a particular residue, or even a whole chain/segment of the macromolecule. For each bead in the model, the frictional force being exerted on the bead by the bulk solvent and perturbed by other beads in the system as it travels through the solvent is calculated. The set of N calculated frictional forces (where N is the number of beads in the model) is arranged into a matrix from which the frictional tensors can be derived and ultimately, the intrinsic viscosity and translational and rotational properties can be calculated. The calculation time is usually in the order of N^3 which can be a challenge for large protein systems but is generally not a problem for smaller system, such as G-quadruplexes.

The first study of G-quadruplex structures in solution using hydrodynamic bead modeling was reported in 1999 by Niermann *et al.* [93]. In that study, the rotational and translational diffusion coefficients were calculated for the Watson–Crick double helical B-DNA, the single-stranded duplex "hairpin," and the four-stranded tetramolecular G-quadruplex structures. The findings demonstrated good agreement between the calculated and experimental hydrodynamic properties when proper hydration of the molecules was considered. In 2005, Li *et al.* employed hydrodynamic bead modeling to study unimolecular G-quadruplex structures [54]. In that study the program HYDROPRO was used to calculate the sedimentation coefficient ($S_{20,w}$) and the translational diffusion coefficient (D_t) for the human telomeric sequence $dAG_3(T_2AG_3)_3$ from the reported high-resolution NMR structure in Na^+ (143D) [52] and the crystal structure in K^+ (1KF1) [53]. HYDROPRO is a program developed by García de la Torre and colleagues for the calculation of hydrodynamic properties of a macromolecule of arbitrary shape. In contrast to its predecessor HYDRO [94], which requires the input of a premade bead model, HYDROPRO can read in an atomic coordinate file (e.g., PDB formatted file) and generate a primary hydrodynamic bead model for calculations. To generate this bead model, HYDROPRO replaces the non-hydrogen atoms with a collection of overlapping spheres (Fig. 2). HYDROPRO then calculates the hydrodynamic

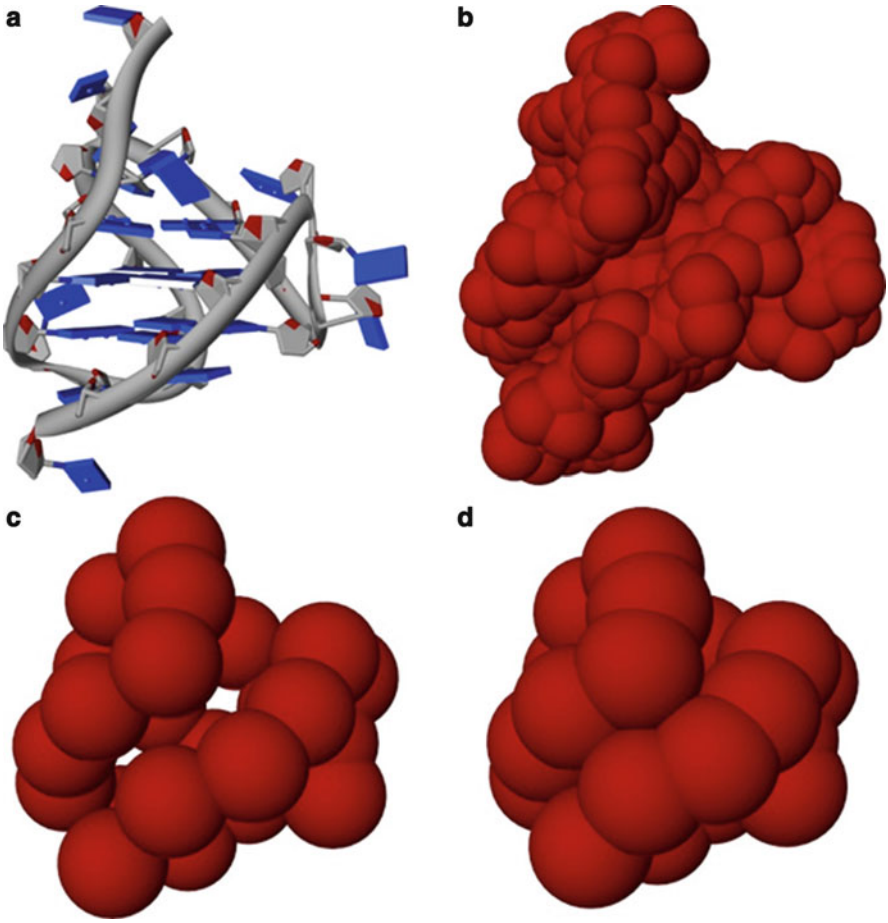


Fig. 2 HYDROPRO generated bead models. The G-quadruplex structure (PDB: 2JPZ) formed from the human telomere sequence, $d(T_2AG_3)T_2$, is shown as a sample of the structures employed for the HYDROPRO calculations. The G-quadruplex is represented (**a**) as a ribbon model with *rectangular blocks* representing the bases. The primary hydrodynamic model is displayed for (**b**) atomic-level shell-model calculation mode, (**c**) residue-level shell-model calculation mode, and (**d**) residue-level bead-model calculation mode

properties of this primary model using the shell-model methodology [95] proposed by Bloomfield *et al*., in which the molecular surface of the model is replaced by a shell of smaller overlapping spheres. The theory behind the shell-model method is that only the surface exposed to the bulk solvent contributes to the frictional properties and thus the hydrodynamic behavior of the molecules.

Using HYDROPRO, it was found that using the Na^+ or "basket" structure experimentally determined by NMR yielded calculated $S_{20,w}$ and D_t values that agreed with the experimentally obtained hydrodynamic measurements via analytical

ultracentrifugation (3.2% lower and 6.0% lower, respectively). However, using the K^+ or "propeller" structure experimentally determined by crystallography yielded calculated $S_{20,w}$ and D_t values that were significantly lower than and inconsistent with the experimentally obtained hydrodynamic measurements (24.9% lower and 18.5% lower, respectively). HYDROPRO was also employed to calculate the $S_{20,w}$-values for a series of *in silico* models to determine possible consistent topologies for the human telomeric sequence in K^+ solution. However, none of the topologies examined (seven antiparallel, one parallel, and one mixed) yielded a calculated $S_{20,w}$-value sufficiently close enough to the experimentally determined $S_{20,w}$-value measurement. It should be noted that the hybrid topologies were unknown at that time and, with their discovery later, the calculated $S_{20,w}$-values of these structures agreed with the experimentally obtained $S_{20,w}$-value for the human telomeric sequence in K^+ solution.

Between 2008 and 2011, hydrodynamic calculations by HYDROPRO were used by Petraccone *et al.* in a novel strategy to study higher-order G-quadruplex structures formed by the human telomeric sequence $(T_2AG_3)_nT_2$ ($n = 4$, 8, 12) [96–98]. Since the 3′-single-stranded overhang of the telomere can be up to 200 bases long, it follows that multiple G-quadruplexes must form on one telomeric overhang. High-resolution monomeric G-quadruplex hybrid topology structures (2HY9, 2JPZ, and 1KF1) were employed as building blocks to construct five different higher-order G-quadruplex models. The five models were hybrid-11 (5′-hybrid-1-hybrid-1-3′), hybrid-12 (5′-hybrid-1-hybrid-2-3′), hybrid-21 (5′-hybrid-2-hybrid-1-3′), hybrid-22 (5′-hybrid-2-hybrid-2-3′), and all-propeller (5′-parallel-parallel-3′). Following molecular dynamics simulations, $S_{20,w}$-value and solvent accessible surface areas (SASA) were calculated using HYDROPRO and NACCESS, respectively, for 400 snapshots over the last 4 ns of simulation. The calculated $S_{20,w}$-values were compared with sedimentation velocity experiments while the SASA-values were compared with 2-aminopurine fluorescent experiments [91]. Only one model (hybrid-12) emerged where both the experimentally determined $S_{20,w}$ (2.3% lower) and SASA-value matched the predicted values and the authors concluded that this structure is the most probable structure in solution [96]. Two models (hybrid-21 and all-propeller) yielded calculated $S_{20,w}$ that were as close or even closer in agreement with the experimentally obtained $S_{20,w}$. However, the predicted SASA-values for these two models did not match the experimentally measured SASA-values. Conversely, one model (hybrid-22) yielded a predicted SASA-value that appeared to match the experimentally determined values; however, the $S_{20,w}$ was not as close as the hybrid-12 model. The findings reported by Petraccone *et al.* illustrate one of the limitations of hydrodynamic calculations in that it is a low resolution technique that should not be employed alone. However, when used in conjunction with other techniques, it could be a powerful tool for predicting possible high-resolution structures. In a follow-up study looking at trimeric G-quadruplex models, it was determined that the structure for the trimeric G-quadruplex is consistent with the hydrodynamic calculations and experimentally derived properties and follows the hybrid-121 model [97]. Higher-order G-quadruplex structures are discussed in more detail in another chapter in this volume.

3.2 Calibration of HYDROPRO Parameters

Two major releases of HYDROPRO have been made publicly available for download (releases 5 and 10). The first versions (5–7) only supported one calculation mode consisting of an atomic-level primary model and shell-model calculation [99]. The latest release (version 10) offers several new enhanced capabilities and features three different calculation modes – atomic-level primary model with shell-model calculation, residue-level primary model with shell-model calculation, and residue-level primary model with bead-model calculation – as well as some support for parallel computing, which offers a great improvement in computing time [100].

Following the procedures outlined by García de la Torre [100], we calibrated the HYDROPRO parameters for use in calculating the hydrodynamic properties of G-quadruplex structures. D_t and $S_{20,w}$ were calculated for nine different structures (Table 1) and compared to the values obtained using sedimentation velocity experiments (Table 2). From these nine structures, seven were used to optimize the HYDROPRO parameters. For the G-quadruplex structures used in this study, the NMR solutions were reported as ensembles of 6–12 structures. HYDROPRO calculations were carried out for all structures in the ensembles and the values from the different reported structures were averaged to obtain the calculated value. To determine the experimental value for a particular G-quadruplex structure, the raw sedimentation data were analyzed by two different software packages, DCDT+ (version 2.3.2, John Philo, Thousand Oaks, CA) and Sedfit (www.analyticalultracentrifugation.com). The experimental value is the average of the two values determined by the programs from the analytical ultracentrifugation data. For both experimental and calculated values, the 95% confidence limit is reported as the error (Fig. 3). In order to determine the optimum atomic element radius (AER) for the beads used in the primary hydrodynamic model for each calculation mode, we fitted the calculated values to the experimental values using a global-fit approach (1) described in the latest HYDROPRO calibration report [100] where the differences between the calculated and experimental values are averaged, first, over all the hydrodynamic properties available for each G-quadruplex structure and then over the entire set of all G-quadruplexes:

$$\Delta^2 = \frac{1}{N_{\mathrm{G-quadruplex}}} \sum^{N_{\mathrm{G-quadruplex}}} \frac{1}{N_{\mathrm{property}}} \sum_{X}^{N_{\mathrm{property}}} \left[\frac{\mathrm{calculated}_X - \mathrm{experimental}_X}{\mathrm{experimental}_X} \right]^2 \quad (1)$$

The Δ value is the root mean-square relative difference between the calculated values and the experimental values with 100Δ representing the percent difference typically used to characterize the goodness of prediction. To evaluate the goodness of prediction for a particular hydrodynamic property, a reduced equation (2) can be used and the $100\Delta_X$ is the percent difference between calculated and experimental values for a particular hydrodynamic property X:

Table 1 G-quadruplex-forming sequences for HYDROPRO calculations

Sequence	Buffer condition	Corresponding structure (PDB ID)
1. $dAG_3(T_2AG_3)_3$	10 mM tetrabutylammoniumphosphate monobasic, 400 mM sodium chloride, pH 7.0	143D
2. $dAG_3(T_2AG_3)_3$	10 mM tetrabutylammoniumphosphate monobasic, 400 mM potassium chloride, pH 7.0	1KF1
3. $dT_2G_3(T_2AG_3)_3A$	10 mM tetrabutylammoniumphosphate monobasic, 400 mM potassium chloride, pH 7.0	2GKU
4. $dA_3G_3(T_2AG_3)_3A_2$	10 mM tetrabutylammoniumphosphate monobasic, 400 mM potassium chloride, pH 7.0	2HY9
5. $dT_2AG_3(T_2AG_3)_3T_2$	10 mM tetrabutylammoniumphosphate monobasic, 400 mM potassium chloride, pH 7.0	2JPZ
6. $dTAG_3(T2AG3)_3T_2$	10 mM tetrabutylammoniumphosphate monobasic, 400 mM potassium chloride, pH 7.0	2JSL
7. $dTAG_3(T_2AG_3)_3$	10 mM tetrabutylammoniumphosphate monobasic, 400 mM potassium chloride, pH 7.0	2JSM
8. $dG_3(T_2AG_3)_3T$	10 mM tetrabutylammoniumphosphate monobasic, 400 mM potassium chloride, pH 7.0	2KF8
9. $dAG_3(T_2AG_3)_3T$	10 mM tetrabutylammoniumphosphate monobasic, 400 mM potassium chloride, pH 7.0	2KKA

Table 2 Summary of HYDROPRO calculated hydrodynamics values

G-quadruplex (PDB ID)	Experimental[a]	Atomic, shell model	Residue/Cα, shell model	Residue/Cα, bead model
143D	$D_t = 1.32$	1.42	1.44	1.43
	$S_{20,w} = 1.86$	1.84	1.87	1.86
2GKU	$D_t = 1.40$	1.40	1.42	1.41
	$S_{20,w} = 2.00$	1.98	2.01	2.00
2HY9	$D_t = 1.38$	1.41	1.39	1.40
	$S_{20,w} = 2.10$	2.17	2.14	2.15
2JPZ	$D_t = 1.40$	1.38	1.38	1.38
	$S_{20,w} = 2.11$	2.11	2.11	2.11
2JSL	$D_t = 1.43$	1.38	1.38	1.38
	$S_{20,w} = 2.07$	2.02	2.03	2.03
2JSM	$D_t = 1.48$	1.43	1.44	1.43
	$S_{20,w} = 1.95$	1.94	1.96	1.94
2KF8	$D_t = 1.50$	1.48	1.46	1.45
	$S_{20,w} = 1.94$	1.91	1.88	1.88
1KF1	$D_t = 1.47$	1.32	1.38	1.36
	$S_{20,w} = 1.96$	1.71	1.79	1.77
2KKA	$D_t = 1.56$	1.42	1.43	1.41
	$S_{20,w} = 2.00$	1.92	1.93	1.91

[a] D_t is expressed in units of 10^{-6} cm^2/s; $S_{20,w}$ in units of 10^{-13} s; 1KF1 and 2KKA were not used in global analysis of best fit AER for primary hydrodynamic model

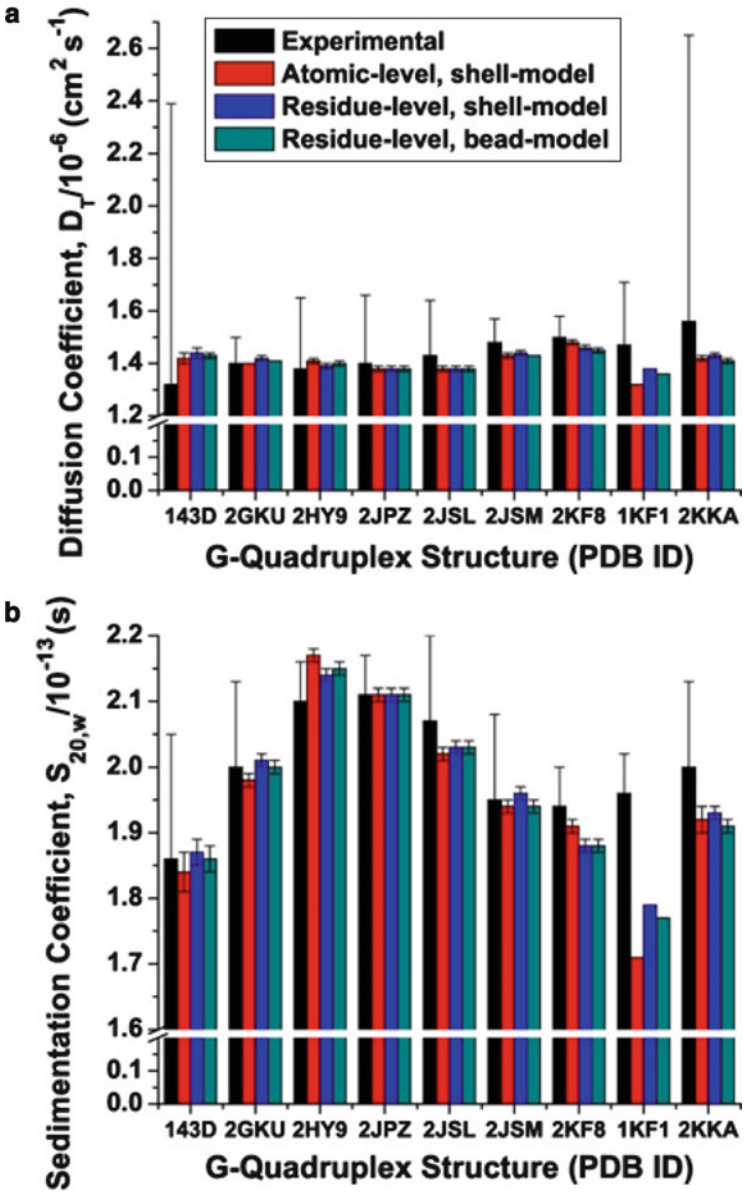


Fig. 3 Summary of HYDROPRO calculations. The translational diffusion coefficients (**a**) and sedimentation coefficients (**b**) for the G-quadruplex structures formed from the human telomere sequence calculated by HYDROPRO using the indicated calculation method and compared to the experimental values. The *error bars* represent the 95% confidence interval. For experimental values, the errors represent the difference between experimental values determined by the DCDT+ software package compared to the Sedfit software package. For calculated values, the errors represent the difference between values calculated from the different solution structures reported. For 1KF1, there was no error for the calculated value since only one crystal structure was reported

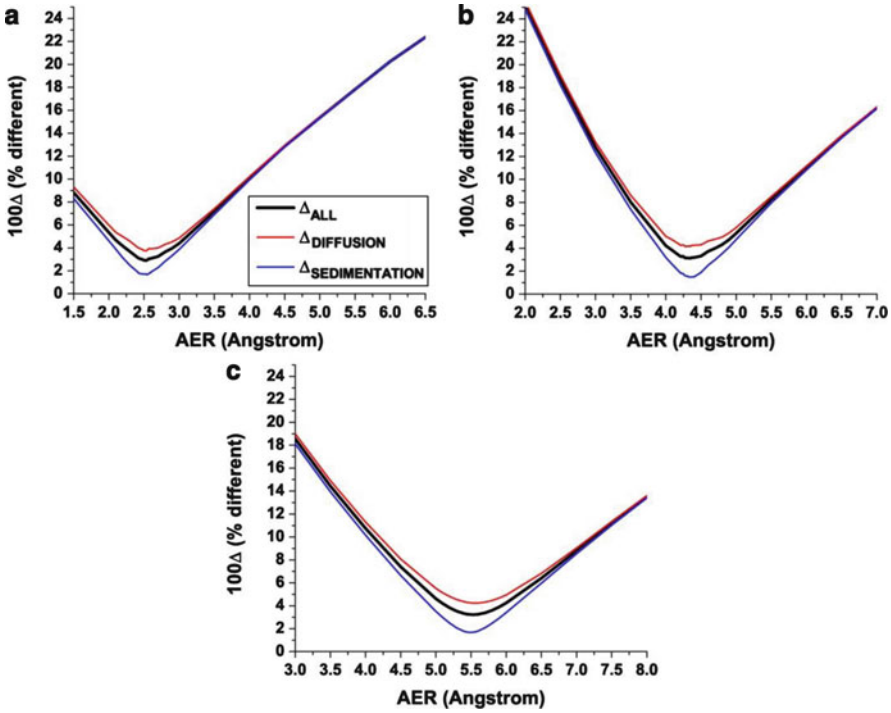


Fig. 4 Global fitting of HYDROPRO parameters. The values for $100\Delta_X$ and 100Δ as a function of the atomic element radius (AER) for the primary hydrodynamic model as calculated using the seven G-quadruplex structures formed from the human telomere sequence. Hydrodynamic properties of G-quadruplexes were calculated using the atomic-level hydrodynamic shell-model calculation (**a**), residue-level shell-model calculation (**b**), and residue-level bead-model calculation (**c**)

$$\Delta_X^2 = \frac{1}{N_{\mathrm{G-quadruplex}}} \sum_{X}^{N_{\mathrm{G-quadruplex}}} \left[\frac{\mathrm{calculated}_X - \mathrm{experimental}_X}{\mathrm{experimental}_X} \right]^2 \tag{2}$$

From the original set of nine structures, the calculated hydrodynamic values for seven NMR solution structures were selected in the global fit analysis (Fig. 4). The PDB IDs for the seven structures were 143D, 2GKU, 2HY9, 2JPZ, 2JSL, 2JSM, and 2KF8. These seven structures were selected because they are solution structures and the sequences contained only canonical bases. For the two structures that were excluded from the global fit analysis (i.e., 1KF1, the crystal structure, and 2KKA, which contains a non-canonical inosine substituted for guanine), the hydrodynamic values were still calculated by HYDROPRO and compared with the experimental values. The results for these two structures are discussed in a subsequent section. For the atomic-level primary hydrodynamic model with HYDROPRO shell-model calculations, the best-fit AER was determined to be

Table 3 Results of HYDROPRO global fit analysis

Calculation mode	AER (Å)	% Diff.	% Diff. $S_{20,w}$	% Diff. D
Atomic, shell model	2.53	2.90	1.71	3.73
Atomic, shell model	2.84	3.77	3.02	4.39
Residue/Cα, shell model	4.35	3.12	1.47	4.16
Residue/Cα, shell model	4.84	4.44	3.74	5.04
Residue/Cα, bead model	5.54	3.22	1.70	4.23
Residue/Cα, bead model	6.11	4.67	3.96	5.29

Table 4 Results of hydration analysis

Calculation mode	AER (Å)	Hydration (g/g)
Atomic, shell model	2.53	0.27
Atomic, shell model	2.84	0.38
Residue/Cα, shell model	4.35	0.04
Residue/Cα, shell model	4.84	0.21
Residue/Cα, bead model	5.54	0.36
Residue/Cα, bead model	6.11	0.56

2.53 Å with 100Δ to be 2.90 (Table 3). This can be interpreted to be that hydrodynamic values predicted at this AER using this particular calculation method have an error of about 2.90% compared to the experimental values. For a non-hydrogen atom, the atomic radius is approximately 1.8 Å [100]. The determined best fit AER suggests a hydration sphere of about 0.7 Å which is slightly less than the 1.1 Å typically used in hydrodynamic calculations to estimate hydration [101]. This smaller hydration sphere represents a hydration of about 0.27 g/g (Table 4). For comparison, the typical hydration value for nucleic acids is 0.35 g/g [102]. The best fit AER determined for G-quadruplex was in close agreement with the standard 2.84 Å AER that was the default setting. When the default AER was used instead of the best fit AER, the error increased from 2.90% to 3.77% (Table 1) of which the major contribution was due to the increased error in predicting the sedimentation coefficient. In the current release of HYDROPRO (version 10), there is an option for calculating hydrodynamic properties at the residue level rather than the atomic level with the bead representation centered on the Cα carbon for each residue. We performed the global fit analysis for the residue-level primary hydrodynamic model with a HYDROPRO shell-model calculation and found that the best fit AER at this level was 4.35 Å with an error of 3.12%. For the residue-level primary hydrodynamic model with a HYDROPRO bead-model calculation, the best fit AER was 5.54 Å with an error of 3.22%. For both of these calculation modes, the increased error compared to the atomic-level calculation was mainly due to the worse prediction of the diffusion translational coefficient. The best fit AER reported for these two modes were similar to the default AERs, 4.84 Å for the residue-level shell-model calculation (5.04% error) and 6.11 Å for the residue-level bead-model calculation (5.29% error). We should note that in all three calculation modes, the best fit AER for each individual hydrodynamic property was in very close agreement with the best fit AER for the entire model (Fig. 4).

All three modes of calculations were able to predict accurately the hydrodynamic properties of the seven G-quadruplex structures fitted (Table 2). The errors of the prediction were well within the range of experimental error for determining the values of hydrodynamic properties, which can be about 5% or more (Fig. 3) [102]. Of the nine original structures considered in this study, two were excluded from the global fit analysis. One of these, 1KF1, was not used because this structure has been previously demonstrated not to predict accurately the hydrodynamic behavior of the structure in solution [54]. Our current findings agree with the previously reported data (Table 2). The other structure, 2KKA, was excluded because the reported structure contained a non-canonical inosine substituted for a guanine. Inosine disrupts G-quadruplex formation [70] and is used to reduce the number of possible folding topologies for a polymorphic G-quadruplex sequence. Instead of using this structure to build our global fit model, we selected it to study how base substitution affects the structure in solution. We found that the calculated hydrodynamic values for 2KKA differ from the experimental values with errors of 7.16%, 6.66%, and 7.47% depending on the calculation method used. Compared with the training set of the seven structures, the error associated with the 2KKA structure is nearly doubled. The findings indicate that, similar to 1KF1, the structure of 2KKA that is selected by the inosine substitution might not be representative of the original structure of the sequence in solution.

4 Limitations of Hydrodynamic Bead Modeling

4.1 Hydrodynamic Bead Modeling is a Low Resolution Technique

Among the chief limitations to hydrodynamic bead modeling is that hydrodynamic measurement itself is a low-resolution technique. The model that accurately predicts the hydrodynamic parameters measured in solution is not necessarily the definitive structure for that molecule but rather one of the possible conformations amongst many others [1]. For example, in the previous study by Petraccone *et al.* [96] where hydrodynamic modeling was employed to predict the structures for higher-order G-quadruplexes, three out of the five models examined agreed with the measured hydrodynamic values. However, only one of those three models agreed with the prediction for solvent accessibility for the adenine bases compared to steady-state fluorescent measurements, suggesting that it was the consistent model. Hydrodynamic measurements can help rule out a particular conformation if the predicted values differ greatly with the experimental values. However, when that difference is smaller, hydrodynamic predictions should be confirmed by non-hydrodynamic experiments (e.g., 2-aminopurine fluorescence spectroscopy to probe the solvent accessibility of adenine bases or DMS footprinting to probe guanine base-paring interactions). Such is the case with the antiparallel human telomeric G-quadruplex structure, 2KKA. To obtain the enrichment needed for

NMR structure determination, an inosine was substituted for the guanine 14 in the original sequence (G14I). The reported NMR structure thus contained this substitution and, from our calculations and measurements on the parent unsubstituted sequence, we can conclude that the structure selected by inosine substitution appears to differ from the structure or ensemble of structures in solution (without inosine substitution) because of the relatively large error associated with the predicted hydrodynamic properties for this structure (~7%), almost double the error seen with the predictions from the global fit analysis (Table 3). However, such an error of prediction still falls within the experimental error associated with hydrodynamic measurement, which can be 5% or greater [102]. Therefore we cannot conclude as to the reason for this apparent discrepancy if it is due to the enrichment of a small subset of topologies that lead to altered hydrodynamic behavior or whether it is because a new structure was populated. Perhaps in this case, a similar approach to probe the loop structure by 2-aminopurine fluorescent measurement is warranted to provide additional insight into the structures of the of 2KKA G-quadruplex.

4.2 Treatment of Hydration in Hydrodynamic Bead Modeling

The structure of a macromolecule in solution, its size, shape, and hydrodynamic properties, can be influenced by the bulk solvent surrounding it through the process of hydration. The accepted interpretation for hydration is that any water whose residence time on the surface of a macromolecule is longer than the hydrodynamic relaxation time (~10 ns) is considered to be a part of the macromolecule and contributes to it hydrodynamic values. There are two current hypotheses to the shape water molecules assume around the macromolecule. In the first hypothesis, uniform expansion, the size of the macromolecule is assumed to expand by a constant factor in all directions. This uniform expansion hypothesis works well for globular structures but breaks down with elongated structures and can lead to abnormal changes in the predicted properties. The inadequacy of the uniform expansion hypothesis has led to an alternative view of hydration being proposed. Instead of assuming a uniform expansion effect, this second hypothesis assumes that hydration forms a layer of uniform thickness around the macromolecule. However, this uniform thickness can result in a relatively thicker hydration layer in the directions where the macromolecule is longer. In HYDROPRO, by replacing the atoms with a representation of beads with a slightly larger size, the effect of hydration on the hydrodynamic properties of the macromolecule can be implicitly accounted for by this uniform hydration hypothesis. The degree of hydration, δ, in units of g/g (grams of water bound per gram of dry solute), can be described as a ratio between the hydrated volume, V_{hyd}, and the anhydrous volume, V_{anh}. The anhydrous volume for a DNA or RNA molecule can be calculated using

$$V_{\text{anhydrous}} = \frac{M\bar{\upsilon}}{N_A} \tag{3}$$

M is the molecular weight, $\bar{\upsilon}$ is the partial specific volume, and N_A is Avogadro's number. For G-quadruplex DNA structures, the partial specific volume is assumed to be 0.55 mL/g although there can be small variations from structure to structure [103]. The degree of hydration, δ, is then calculated using

$$\delta = \left(\frac{V_{\text{hydrated}}}{V_{\text{anhydrous}}} - 1\right)\bar{\upsilon}\rho \tag{4}$$

Using (3) and (4) and the hydrated volume of the G-quadruplexes determined by HYDROPRO, we determined the degree of hydration for the seven G-quadruplex structures employed previously in the global fit analysis (i.e., 143D, 2GKU, 2HY9, 2JPZ, 2JSL, 2JSM, 2KF8). The results are shown in Table 4. We did not calculate the degree of hydration for 1KF1 and 2KKA since these reported structures might not be representative of the structures that exist in solution. Using the best-fit AER for each calculation mode, we obtained a degree of hydration of 0.27 g/g for calculations using the atomic-level model with the shell-model calculation, 0.04 g/g for calculations using the residue-level model with the shell-model calculation, and 0.36 g/g for calculations using the residue-level model with the bead-model calculation. The 0.36 g/g degree of hydration predicted by the residue-level model with bead model calculation is the closest to the value of 0.35 g/g reported for nucleic acids by Fernandes *et al.* [102]. However, it should be noted that the nucleic acid sequences examined in that particular study were short sequences whose shapes can be more elongated than the G-quadruplex structures. Thus, it is not inconceivable that the hydration for G-quadruplexes would fall closer to the typical 0.3 g/g for globular proteins [101]. In that case, the 0.27 g/g estimate by the atomic-level model with shell-model calculations is more appropriate. We also performed calculations using the default AER for HYDROPRO and found that the degrees of hydrations were 0.38, 0.21, and 0.56 g/g, respectively, for the three calculation modes. The atomic-level model with shell-model calculation method is the only one with a reasonable estimate of hydration (0.38 g/g). The hydration is overestimated by the residue-level model with bead-model calculation (0.56 g/g) while the residue-level model with shell-model calculation still underestimates it (0.21 g/g). It should be noted that the volume calculated by HYDROPRO is based entirely on the size of the beads in the primary hydrodynamic model and is not dependent on the shell-model/bead-model calculation of frictional properties. Thus, the reason that the residue-level model with shell-model calculation underestimate the size of the G-quadruplexes compared to the residue-level model with bead-model calculation is because the former employs a smaller AER value. In fact, when we used 4.35 Å instead of 5.54 Å for the residue-level model with bead-model calculation, we obtained the same degree of hydration of 0.04 g/g. We can conclude then that the residue-level with shell-model calculation is not an appropriate method for hydrodynamic calculations of G-quadruplex structures.

4.3 Potassium Binding by G-Quadruplex and Its Implication on Hydrodynamics

If water binding can potentially influence the hydrodynamic properties of a G-quadruplex structure, it follows that the same could be said about the binding of potassium and other ions. Traditionally, the association of cations with a G-quadruplex structure is believed to occur primarily within the central ion channel formed by the G-quartets. The coordination of potassium (K^+), sodium (Na^+), thallium (Tl^+), rubidium (Rb^+), cesium (Cs^+), calcium (Ca^{2+}), lead (Pb^{2+}), barium (Ba^{2+}), and strontium (Sr^{3+}) to the G-quadruplex central ion channel has been observed in several crystal structures [104]. However, it was recently reported that, in addition to the canonical internal binding sites, cations can also bind to a G-quadruplex structure externally, particularly within the loop regions [105]. In a study using KCl titration to the human telomere sequence (Tel22 – d[A($G_3T_2A)_3$]), it was determined that K^+ binds to Tel22 at a stoichiometry of $3K^+$/Tel22 at $[K^+]_{free} =$ 5 mM and that stoichiometry increases to 8–10 K^+/Tel22 at $[K^+]_{free} = 20$ mM [106]. In addition, it was reported that the folding of Tel22 into the G-quadruplex structure is driven almost entirely by cation binding with K^+ binding contributing ~−4.9 kcal/mol (at $[K^+]_{free} = 5$ mM) to the overall folding free energy of −2.4 kcal/mol. Together, these findings suggest that, in addition to stabilizing the G-quartets within the G-quadruplex core, potassium ions can play other roles in promoting G-quadruplex formation and stability. This hypothesis is further supported by two structural studies – Phan *et al.* in 2011 that reported a structure of RNA G-quadruplex/duplex junction determined by NMR spectroscopy which was the first structure to demonstrate a possible cation binding site on the surface of a G-quadruplex structure [107] and Wei *et al.* in 2012 which reported a structure of the human *c-Kit* DNA promoter sequence determined by X-ray crystallography with two K^+ and one Mg^{2+} ions binding in the G-quadruplex loops and grooves in addition to the K^+ ions within canonical central ion channel [108]. In the crystal structure, all three external cations are believed to play a role in maintaining the *c-Kit G*-quadruplex structure. It should be noted that one of the external K^+ ions appeared to be transient and capable of adopting one of several distinct positions, while the other external K^+ ion appeared to be more static, suggesting the existence of a high affinity binding site. While the K^+ ions are believed to have a primary role in stabilizing the G-quadruplex structure by direct electrostatic interaction with the DNA, the Mg^{2+} ion is thought to assume a secondary role of shielding the anionic charge of the phosphate groups. In fact, it is well known that polyanions, e.g., DNA, can attract a shell of cations to help partially neutralize the negatively-charged backbone phosphates [109, 110].

To examine how K^+ ions can affect the structure of a G-quadruplex-forming sequence hydrodynamically, we conducted sedimentation velocity experiments for two G-quadruplex-forming sequences in the presence of different concentrations of K^+ ions (Fig. 5). For both sequences an increase in sedimentation coefficient was observed with the corresponding increase in the amount of potassium present. In low potassium solutions (25 mM) the low sedimentation coefficients can

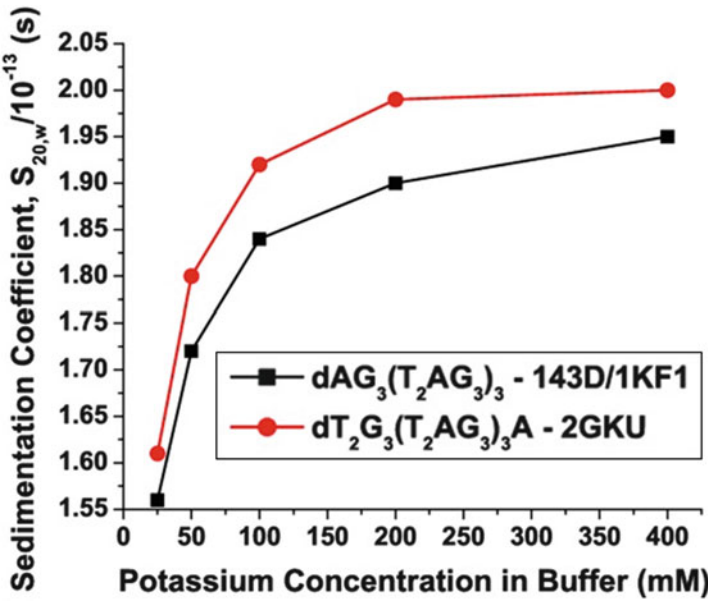


Fig. 5 The effect of potassium on G-quadruplex hydrodynamic measurements. The experimentally determined value for the sedimentation coefficient is dependent on the amount of potassium in solution. This dependency was demonstrated for two different sequences $dAG_3(T_2AG_3)_3$, the same sequence used to derive an X-ray crystal structure in K^+ (1KF1) and the NMR solution structure in Na^+ (143D), and $dT_2G_3(T_2AG_3)A$, the sequence used to derive an NMR solution structure in K^+ (2GKU)

probably be attributed to hydrodynamic non-ideality. In an ideal solution of uncharged particles, the equilibrium distribution of each individual species subjected to a centrifugation field traveling through an incompressible solvent is independent of one another. However, for charged macromolecules such as DNA and some proteins, their equilibrium distributions can be coupled to the distribution of other charged particles (i.e., counter ions such as potassium in our case) by the Donnan effect, resulting in non-ideal behavior [111]. Non-ideality can cause a macromolecule to appear much smaller than it actually is and, in effect, results in a lower molecular weight and sedimentation coefficient from sedimentation experiments. The Donnan effect can be corrected in one of two ways. The first option is to reduce the charge of the macromolecule by adjusting the pH but this option is only applicable to zwitterionic species. The second option is to "swamp" out the Donnan effect by adding a large quantity of a charged electrolytes (i.e., potassium) to the solution. However, at the higher concentration of potassium it is unclear how much of the increase in the sedimentation coefficient can be attributed to a shift toward an ideal solution and how much can be attributed to possible direct potassium binding. The binding of a potassium ion to a G-quadruplex structure can have one of two possible consequences. In the first scenario, the potassium ions are bound to the G-quadruplex on the surface, possibly along the phosphate backbones.

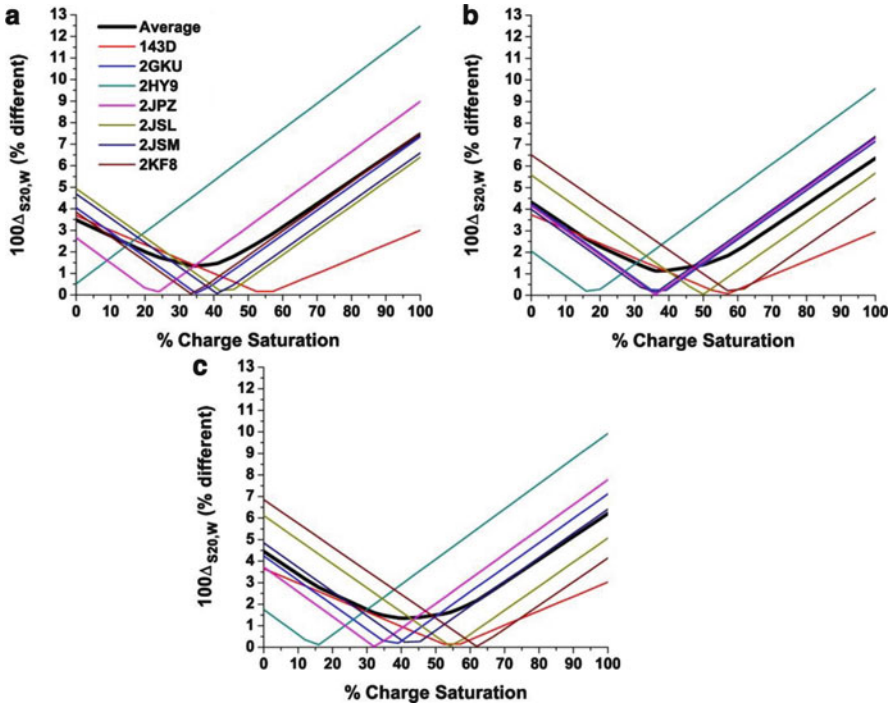


Fig. 6 Results of HYDROPRO parameters calibration by number of potassium bound. The values for $100\Delta S_{20,W}$ as a function of percent charge saturation for the primary hydrodynamic model calculated using the seven G-quadruplex structures formed from the human telomere sequence. Hydrodynamic properties of G-quadruplexes were calculated using atomic-level shell-model calculation (**a**), residue-level shell-model calculation (**b**), and residue-level bead-model calculation (**c**)

In this case, these ions will not only change the apparent molecular weight of the G-quadruplex structures but also its shape. In the second scenario, the potassium ion is trapped within the G-quartet core or is buried deep within one of the grooves. Because these potassium ions are not at the surface of the G-quadruplex structure, they do not change the shape of the hydrodynamic model; however, they would cause an underestimation of the apparent molecular weight leading to an under-prediction of the sedimentation coefficient. To explore this scenario further, we performed a series of HYDROPRO calculations and added potassium or sodium to the G-quadruplex structure by making the appropriate increase in the molecular weight value in the HYDROPRO parameter file. The calculations were done using all three modes of HYDROPRO calculation and the default AER instead of the best-fit AER determined in the previous section 3.2. In order to display the data for the sequences on the same scale, we normalized the number of potassium ions bound by (5) as follow:

$$\%\ \text{charge saturation} = \frac{\#\ \text{potassium bound}}{\text{sequence length (bp)} - 1} \tag{5}$$

The results of the HYDROPRO calculations are shown in Fig. 6. For all three calculation modes, the smallest errors for many of the sequences were observed at 40% charge saturation which is equal to approximately eight to ten bound potassium ions. It should be noted that the correction for molecular weight is as good as or even better than the previous correction for size of the beads (Fig. 4). These findings would suggest that potassium ions not only play a role in G-quadruplex formation and stability but can also influence how the G-quadruplex structure appears in hydrodynamic observations. However, these results require further investigation and validation as there is yet no structural study to back up this hypothesis.

5 Future Directions for G-Quadruplex Hydrodynamic Bead Modeling

G-quadruplex structures have emerged as attractive drug targets because of the overrepresentation of putative quadruplex-forming sequences in functionally important sites within the human genome [17]. In addition, because of their highly polymorphic nature, G-quadruplex-targeted therapies have the potential to be highly selective compared to other DNA targets. However, because of this polymorphism, it can sometimes be difficult to elucidate the relevant structure for a putative G-quadruplex-forming sequence. Modification of the sequence by base subtraction, addition, and substitution or changing of the experimental conditions has been shown to reduce the polymorphism in some cases. However such perturbation can have an unwanted effect by shifting the equilibrium to favor species that might not be indicative of, or be completely dissimilar to, any species actually formed *in vitro* or *in vivo* [62]. The rational drug design of new small molecule G-quadruplex binding agents can be limited by the lack of a definitive target structure. An example of this is the cationic porphyrin, TMPyP4, a G-quadruplex binding agent whose mode of binding and mechanism of action remains a current topic of investigation [82, 112–115]. Hydrodynamic bead modeling presents a potentially useful tool to help address this problem. As seen in the reports by Petraccone *et al.*, measurement of hydrodynamic properties, combined with *in silico* simulations (e.g., molecular dynamics) and comprehensive biophysical characterization could be used to predict new structures that might serve as a basis for rational drug design.

We present a case study on how this can be conducted. The sequence for the G-quadruplex hybrid-1 structure 2HY9 was obtained from Integrated DNA Technologies (Corralville, IA) and consists of the sequence

5′-dAAAGGGTTAGGGTTAGGGTTAGGGAA-3′

A stock solution of the oligonucleotide was dissolved in folding buffer, which is composed of monobasic tetrabutylammoniumphosphate (10 mM), EDTA disodium

salt dihydrate (1 mM), pH 7.0. Samples were made to an optical density of 0.80 A_{260} by dilution with folding buffer. KCl was added to bring the final concentration to 400 mM. The oligonucleotide samples were annealed in a water bath by heating to 100°C, holding the samples at that temperature for 10 min, and gradually cooling to room temperature overnight. Sedimentation velocity experiments were carried out in a Beckman Coulter ProteomeLab XL-A analytical ultracentrifuge at 20°C overnight at 50,000 rpm in standard two sector cells. Data were analyzed using two different software packages, DCDT+ (version 2.3.2, John Philo, Thousand Oaks, CA) and Sedfit (www.analyticalultracentrifugation.com). Excellent agreement between the two software packages for $S_{20,w}$ values and species distributions was achieved. Buffer density was determined on a Mettler/Paar Calculating Density Meter DMA 55A at 20.0°C and buffer viscosity was calculated using Sednterp software (www.jphilo.mailway.com).

The model for a hybrid-1 G-quadruplex structure was downloaded from the Protein Bank Database (PDB ID: 2HY9) and to the first structure in the file, two potassium ions were added between the G-quartets to complete the model. In a different study with a novel G-quadruplex-forming sequence, this structure could be generated using a fragment-based approach [116]. A solvation model with explicit water and counter ions representations were generated using the parm99SB.dat Amber force field with the following protocol: (1) 75 potassium ions were added randomly around the G-quadruplex structures using Amber 11 leap rules for counter ions, (2) charges were neutralized by the addition of chloride ions, and (3) the system was solvated by the addition of a rectangular box of TIP3P water (51 × 52 × 46 Å). The final model represents a system consisting of 13 mM G-quadruplex in 1 M KCl. The system was heated slowly and equilibrated for 600 ps using the following protocol: (1) minimize water, (2) 50 ps molecular dynamics (heating to 300 K) holding the DNA fixed (2 kcal/mol/Å), (3) 50 ps molecular dynamics ($T = 300$ K) for density equilibration holding the DNA fixed (2 kcal/mol/Å), and (4) 500 ps molecular dynamics ($T = 300$ K). Production runs of 200 ns were carried out after the final equilibration step to obtain snapshots of 10 ps throughout the trajectory. Simulations were performed in the isothermal isobaric ensemble ($P = 1$ atm., $T = 300$ K). Periodic boundary conditions and the Particle-Mesh-Ewald algorithm were used. A 2.0-fs time step was used with bonds involving in hydrogen atoms frozen using SHAKE. For the equilibration steps and the production steps, molecular dynamics calculations were carried out using AMBER 11 program sander and the cuda version of pmemd, respectively. The trajectories were analyzed using the PTRAJ module in AMBER and visualized using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco. The hydrodynamic properties of the simulated models were computed using HYDROPRO [100] and the best-fit AERs, 2.53, 4.35, 5.54 Å, previously determined by global analysis fitting. The distributions of the sedimentation coefficients were computed over the last 20 ns of the trajectory (2,000 snapshots in the last 20 ns of simulation).

The results are shown in Fig. 7 with the calculated sedimentation coefficient distribution (colored lines) overlaying the experimentally derived sedimentation

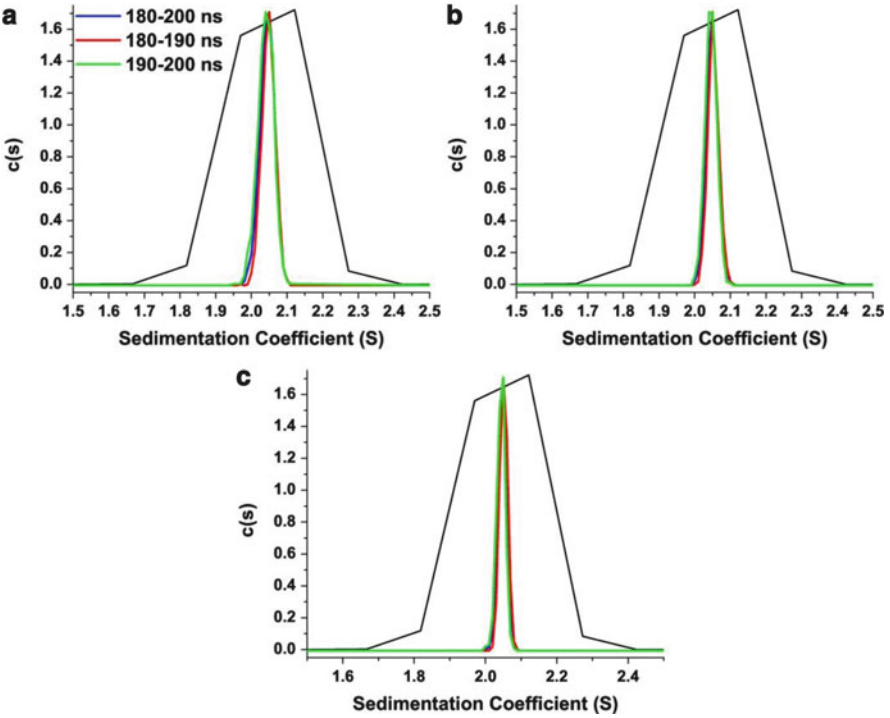


Fig. 7 Results of molecular dynamics and HYDROPRO calculations. The distribution of the sedimentation coefficient, c(s), for the hybrid-1 G-quadruplex structure 2HY9 determined experimentally by sedimentation velocity (*black*) as compared to the distribution of sedimentation coefficient computed by HYDROPRO following molecular dynamics simulation. The sedimentation coefficient was computed for each snapshot taken every 10 ps between the 180–200 ns (*blue*), 180–190 ns (*red*), and 190–200 ns (*green*) interval of the trajectory. Hydrodynamic properties of G-quadruplexes were calculated using atomic-level shell-model calculation (**a**), residue-level shell-model calculation (**b**), and residue-level bead-model calculation (**c**)

coefficient distribution (black line). The results from all three calculation modes – atomic-level model with shell-model calculation, residue-level model with shell-model calculation, and residue-level model with bead-model calculation – were in close agreement with each other. The average of all the sedimentation coefficients from the 2,000 snapshots was considered as the calculated sedimentation coefficient. The sedimentation coefficient was 2.040 for the atomic-level model with the shell-model calculation, 2.044 for the residue-level with the shell-model calculation, and 2.043 for the residue-level model with the bead-model calculation This corresponds to $100\Delta_X$ values of error of 2.65%, 2.41%, and 2.48% for the three calculation modes, respectively. These results agreed with previous calculations performed for the global fit analysis (Table 3). It should be noted that in all cases the experimentally derived sedimentation coefficient distribution was wider than the calculated distribution, possibly due to the larger extent of the experimental error (Fig. 7). In addition, in all three calculation modes, the distribution of sedimentation

coefficient remained unchanged between 180–190 ns and 190–200 ns compared to 180–200 ns, indicating that the simulation had reached equilibrium.

6 Conclusion

We have demonstrated that hydrodynamic bead modeling is a powerful tool for studying G-quadruplex structures when experimentally determined high-resolution structural representations are not available or cannot be feasibly obtained. In such cases, the use of low-resolution techniques such as hydrodynamics, combined with readily accessible biophysical measurements (e.g., CD spectroscopy, fluorescent spectroscopy), can be used to provide a high-resolution glimpse into these structures. One of the limitations of hydrodynamic calculations, such as HYDROPRO, is that the calculation is performed on one structure giving a static look at an otherwise dynamic system. Recent advancements in computation technology, particularly the development of molecular dynamic calculations to take advantage of the speed and calculation power of graphical processor units, allows for the researcher to conduct longer and more complex simulations. These can generate a more dynamic representation of the macromolecule structure that can be validated by hydrodynamic bead modeling and can be used as a basis for drug design. From our work, we recommend for G-quadruplexes using either the atomic-level model with shell-model calculation mode (AER = 2.53 Å) or the residue-level model with bead-model calculation mode (AER = 5.54 Å) for HYDROPRO calculations, as both modes can predict the hydrodynamic properties accurately with a reasonable estimate of the size of the macromolecule. The residue-level model with bead-model calculation mode has the advantage of being significantly faster but with a slightly higher error in predicting hydrodynamic properties. The parameters for HYDROPRO presented here can be used as for hydrodynamic calculation of G-quadruplexes or can be further optimized against a more extensive set of hydrodynamic properties.

Acknowledgments We would like to acknowledge Dr. Robert D. Gray for useful discussions in the writing of this chapter. Molecular graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR001081). This work was supported by NIH Grants CA35635 (J.B.C), GM077422 (J.B.C. and J.O.T) and University of Louisville grant, CTSPGP 20058 Award (J.B.C. and J.O.T).

References

1. Byron O (2008) Hydrodynamic modeling: the solution conformation of macromolecules and their complexes. In: John JC, William Detrich H III (eds) Methods in cell biology, vol 84. Academic Press in Waltham, MA, Burlington

2. Carrasco B, García de la Torre J (1999) Hydrodynamic properties of rigid particles: comparison of different modeling and computational procedures. Biophys J 76(6):3044–3057
3. García de la Torre J, Bloomfield VA (1981) Hydrodynamic properties of complex, rigid, biological macromolecules: theory and applications. Q Rev Biophys 14(01):81–139
4. Hopkins AL, Groom CR (2002) The druggable genome. Nat Rev Drug Discov 1(9):727–730
5. Drews J (2000) Drug discovery: a historical perspective. Science 287(5460):1960–1964
6. Imming P, Sinning C, Meyer A (2006) Drugs, their targets and the nature and number of drug targets. Nat Rev Drug Discov 5(10):821–834
7. Chaires JB (1986) Allosteric conversion of Z DNA to an intercalated right-handed conformation by daunomycin. J Biol Chem 261(19):8899–8907
8. Chaires JB (1986) Inhibition of the thermally driven B to Z transition by intercalating drugs. Biochemistry 25(26):8436–8439
9. Qu X, Trent JO, Fokt I, Priebe W, Chaires JB (2000) Allosteric, chiral-selective drug binding to DNA. Proc Natl Acad Sci USA 97(22):12032–12037
10. Scanlon K (2004) Anti-genes: siRNA, ribozymes and antisense. Curr Pharm Biotechnol 5(5):415–420
11. Sepp-Lorenzino L, Ruddy MK (2008) Challenges and opportunities for local and systemic delivery of siRNA and antisense oligonucleotides. Clin Pharmacol Ther 84(5):628–632
12. de Fougerolles A, Vornlocher H-P, Maraganore J, Lieberman J (2007) Interfering with disease: a progress report on siRNA-based therapeutics. Nat Rev Drug Discov 6(6):443–453
13. Tiemann K, Rossi JJ (2009) RNAi-based therapeutics – current status, challenges and prospects. EMBO Mol Med 1(3):142–151
14. Garzon R, Marcucci G, Croce CM (2010) Targeting microRNAs in cancer: rationale, strategies and challenges. Nat Rev Drug Discov 9(10):775–789
15. Castanotto D, Rossi JJ (2009) The promises and pitfalls of RNA-interference-based therapeutics. Nature 457(7228):426–433
16. Williamson JR, Raghuraman MK, Cech TR (1989) Monovalent cation-induced structure of telomeric DNA: the G-quartet model. Cell 59(5):871–880
17. Huppert JL, Balasubramanian S (2005) Prevalence of quadruplexes in the human genome. Nucleic Acids Res 33(9):2908–2916
18. Huppert JL, Balasubramanian S (2007) G-quadruplexes in promoters throughout the human genome. Nucleic Acids Res 35(2):406–413
19. Neidle S (2010) Human telomeric G-quadruplex: the current status of telomeric G-quadruplexes as therapeutic targets in human cancer. FEBS J 277(5):1118–1125
20. De Cian A, Lacroix L, Douarre C, Temime-Smaali N, Trentesaux C, Riou JF, Mergny JL (2008) Targeting telomeres and telomerase. Biochimie 90(1):131–155
21. Wright WE, Tesmer VM, Huffman KE, Levene SD, Shay JW (1997) Normal human chromosomes have long G-rich telomeric overhangs at one end. Genes Dev 11 (21):2801–2809
22. Riou JF, Guittat L, Mailliet P, Laoui A, Renou E, Petitgenet O, Mégnin-Chanet F, Hélène C, Mergny JL (2002) Cell senescence and telomere shortening induced by a new series of specific G-quadruplex DNA ligands. Proc Natl Acad Sci USA 99(5):2672–2677
23. Cuesta J, Read MA, Neidle S (2003) The design of G-quadruplex ligands as telomerase inhibitors. Mini Rev Med Chem 3(1):11
24. Lopes J, Piazza A, Bermejo R, Kriegsman B, Colosio A, Teulade-Fichou M-P, Foiani M, Nicolas A (2011) G-quadruplex-induced instability during leading-strand replication. EMBO J 30(19):4033–4046
25. Rodriguez R, Miller KM, Forment JV, Bradshaw CR, Nikan M, Britton S, Oelschlaegel T, Xhemalce B, Balasubramanian S, Jackson SP (2012) Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. Nat Chem Biol 8(3):301–310
26. Shay JW, Bacchetti S (1997) A survey of telomerase activity in human cancer. Eur J Cancer 33(5):787–791
27. Ambrus A, Chen D, Dai JX, Jones RA, Yang DZ (2005) Solution structure of the biologically relevant g-quadruplex element in the human *c-MYC* promoter. Implications for g-quadruplex stabilization. Biochemistry 44(6):2048–2058

28. Hsu STD, Varnai P, Bugaut A, Reszka AP, Neidle S, Balasubramanian S (2009) A G-rich sequence within the c-kit oncogene promoter forms a parallel G-quadruplex having asymmetric G-tetrad dynamics. J Am Chem Soc 131(37):13399–13409
29. Phan AT, Kuryavyi V, Burge S, Neidle S, Patel DJ (2007) Structure of an unprecedented G-quadruplex scaffold in the human *c-Kit* promoter. J Am Chem Soc 129(14):4386–4392
30. Dai J, Chen D, Jones RA, Hurley LH, Yang D (2006) NMR solution structure of the major G-quadruplex structure formed in the human *BCL2* promoter region. Nucleic Acids Res 34(18):5133–5144
31. Sun DY, Guo KX, Rusche JJ, Hurley LH (2005) Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human *VEGF* gene by the presence of potassium and G-quadruplex-interactive agents. Nucleic Acids Res 33(18):6070–6080
32. De Armond R, Wood S, Sun DY, Hurley LH, Ebbinghaus SW (2005) Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1 alpha promoter. Biochemistry 44(49):16341–16350
33. Eddy J, Maizels N (2006) Gene function correlates with potential for G4 DNA formation in the human genome. Nucleic Acids Res 34(14):3887–3896
34. Verma A, Halder K, Halder R, Yadav VK, Rawal P, Thakur RK, Mohd F, Sharma A, Chowdhury S (2008) Genome-wide computational and expression analyses reveal G-quadruplex DNA motifs as conserved *cis*-regulatory elements in human and related species. J Med Chem 51(18):5641–5649
35. Brooks TA, Kendrick S, Hurley L (2010) Making sense of G-quadruplex and i-motif functions in oncogene promoters. FEBS J 277(17):3459–3469
36. Balasubramanian S, Hurley LH, Neidle S (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? Nat Rev Drug Discov 10(4):261–275
37. Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress *c-MYC* transcription. Proc Natl Acad Sci USA 99(18):11593–11598
38. Huppert JL, Bugaut A, Kumari S, Balasubramanian S (2008) G-quadruplexes: the beginning and end of UTRs. Nucleic Acids Res 36(19):6260–6268
39. Halder K, Wieland M, Hartig JS (2009) Predictable suppression of gene expression by 5′-UTR-based RNA quadruplexes. Nucleic Acids Res 37(20):6811–6817
40. Arora A, Suess B (2011) An RNA G-quadruplex in the 3′ UTR of the proto-oncogene *PIM1* represses translation. RNA Biol 8(5):802–805
41. Marcel V, Tran PLT, Sagne C, Martel-Planche G, Vaslin L, Teulade-Fichou M-P, Hall J, Mergny J-L, Hainaut P, Van Dyck E (2011) G-quadruplex structures in *TP53* intron 3: role in alternative splicing and in production of p53 mRNA isoforms. Carcinogenesis 32(3):271–278
42. Bang I (1910) Untersuchungen über die Guanylsäure. Biochem Z 26:293–311
43. Gellert M, Lipsett MN, Davies DR (1962) HELIX formation by guanylic acid. Proc Natl Acad Sci USA 48(12):2013–2018
44. Ghosh A, Bansal M (2003) A glossary of DNA structures from A to Z. Acta Crystallogr D 59(4):620–626
45. Laughlan G, Murchie A, Norman D, Moore M, Moody P, Lilley D, Luisi B (1994) The high-resolution crystal structure of a parallel-stranded guanine tetraplex. Science 265(5171): 520–524
46. Phillips K, Dauter Z, Murchie AIH, Lilley DMJ, Luisi B (1997) The crystal structure of a parallel-stranded guanine tetraplex at 0.95 Å resolution. J Mol Biol 273(1):171–182
47. Smith FFWJ (1992) Quadruplex structure of *Oxytricha* telomeric DNA oligonucleotides. Nature 356(6365):164–168
48. Smith FW, Feigon J (1993) Strand orientation in the DNA quadruplex formed from the *Oxytricha* telomere repeat oligonucleotide $d(G_4T_4G_4)$ in solution. Biochemistry 32(33): 8682–8692

49. Haider S, Parkinson GN, Neidle S (2002) Crystal structure of the potassium form of an *Oxytricha nova* G-quadruplex. J Mol Biol 320(2):189–200
50. Horvath MP, Schultz SC (2001) DNA G-quartets in a 1.86 Å resolution structure of an *Oxytricha nova* telomeric protein-DNA complex. J Mol Biol 310(2):367–377
51. Kang C, Zhang X, Ratliff R, Moyzis R, Rich A (1992) Crystal structure of four-stranded *Oxytricha* telomeric DNA. Nature 356(6365):126–131
52. Wang Y, Patel DJ (1993) Solution structure of the human telomeric repeat d[$AG_3(T_2AG_3)_3$] G-tetraplex. Structure 1(4):263–282
53. Parkinson GN, Lee MPH, Neidle S (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. Nature 417(6891):876–880
54. Li J, Correia JJ, Wang L, Trent JO, Chaires JB (2005) Not so crystal clear: the structure of the human telomere G-quadruplex in solution differs from that present in a crystal. Nucleic Acids Res 33(14):4649–4659
55. Miller MC, Buscaglia R, Chaires JB, Lane AN, Trent JO (2010) Hydration is a major determinant of the G-quadruplex stability and conformation of the human telomere 3′ sequence of d[$AG_3(TTAG_3)_3$]. J Am Chem Soc 132(48):17105–17107
56. Heddi B, Phan AT (2011) Structure of human telomeric DNA in crowded solution. J Am Chem Soc 133(25):9824–9833
57. Ambrus A, Chen D, Dai J, Bialis T, Jones RA, Yang D (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. Nucleic Acids Res 34(9):2723–2735
58. Luu KN, Phan AT, Kuryavyi V, Lacroix L, Patel DJ (2006) Structure of the human telomere in K + solution: an intramolecular (3+1) G-quadruplex scaffold. J Am Chem Soc 128 (30):9963–9970
59. Dai J, Carver M, Punchihewa C, Jones RA, Yang D (2007) Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K^+ solution: insights into structure polymorphism of the human telomeric sequence. Nucleic Acids Res 35(15):4927–4940
60. Lim KW, Amrane S, Bouaziz S, Xu W, Mu Y, Patel DJ, Luu KN, Phan AT (2009) Structure of the human telomere in K^+ solution: a stable basket-type G-quadruplex with only two G-tetrad layers. J Am Chem Soc 131(12):4301–4309
61. Schultze P, Macaya RF, Feigon J (1994) Three-dimensional solution structure of the thrombin-binding DNA aptamer d(GGTTGGTGTGGTTGG). J Mol Biol 235(5):1532–1547
62. Lane AN, Chaires JB, Gray RD, Trent JO (2008) Stability and kinetics of G-quadruplex structures. Nucleic Acids Res 36(17):5482–5515
63. Dailey MM, Miller MC, Bates PJ, Lane AN, Trent JO (2010) Resolution and characterization of the structural polymorphism of a single quadruplex-forming sequence. Nucleic Acids Res 38(14):4877–4888
64. Dai J, Carver M, Yang D (2008) Polymorphism of human telomeric quadruplex structures. Biochimie 90(8):1172–1183
65. Yang D, Okamoto K (2010) Structural insights into G-quadruplexes: towards new anticancer drugs. Future Med Chem 2(4):619–646
66. Virgilio A, Esposito V, Randazzo A, Mayol L, Galeone A (2005) 8-Methyl-2′-deoxyguanosine incorporation into parallel DNA quadruplex structures. Nucleic Acids Res 33(19):6188–6195
67. Virgilio A, Esposito V, Randazzo A, Mayol L, Galeone A (2005) Effects of 8-methyl-2′-deoxyadenosine incorporation into quadruplex forming oligodeoxyribonucleotides. Bioorg Med Chem 13(4):1037–1044
68. Esposito V, Randazzo A, Piccialli G, Petraccone L, Giancola C, Mayol L (2004) Effects of an 8-bromodeoxyguanosine incorporation on the parallel quadruplex structure [d(TGGGT)](4). Org Biomol Chem 2(3):313–318
69. Mekmaysy CS, Petraccone L, Garbett NC, Ragazzon PA, Gray RD, Trent JO, Chaires JB (2008) Effect of O6-methylguanine on the stability of G-quadruplex DNA. J Am Chem Soc 130(21):6710–6711

70. Petrovic AG, Polavarapu PL (2008) Quadruplex structure of polyriboinosinic acid: dependence on alkali metal ion concentration, pH and temperature. J Phys Chem B 112(7): 2255–2260
71. Marathias VM, Sawicki MJ, Bolton PH (1999) 6-Thioguanine alters the structure and stability of duplex DNA and inhibits quadruplex DNA formation. Nucleic Acids Res 27 (14):2860–2867
72. Spackova N, Cubero E, Sponer J, Orozco M (2004) Theoretical study of the guanine – > 6-thioguanine substitution in duplexes, triplexes, and tetraplexes. J Am Chem Soc 126 (44):14642–14650
73. Gros J, Avino A, de la Osa JL, Gonzalez C, Lacroix L, Perez A, Orozco M, Eritja R, Mergny JL (2008) 8-Amino guanine accelerates tetramolecular G-quadruplex formation. Chem Commun (25):2926–2928
74. Esposito V, Virgilio A, Randazzo A, Galeone A, Mayol L (2005) A new class of DNA quadruplexes formed by oligodeoxyribonucleotides containing a 3′-3′ or 5′-5′ inversion of polarity site. Chem Commun (31):3953–3955
75. Bonifacio L, Church FC, Jarstfer MB (2008) Effect of locked-nucleic acid on a biologically active G-quadruplex. A structure-activity relationship of the thrombin aptamer. Int J Mol Sci 9(3):422–433
76. Kumar N, Maiti S (2007) Role of locked nucleic acid modified complementary strand in quadruplex/Watson–Crick duplex equilibrium. J Phys Chem B 111(42):12328–12337
77. Tang CF, Shafer RH (2006) Engineering the quadruplex fold: nucleoside conformation determines both folding topology and molecularity in guanine quadruplexes. J Am Chem Soc 128(17):5966–5973
78. Qi J, Shafer RH (2007) Human telomere quadruplex: refolding and selection of individual conformers via RNA/DNA chimeric editing. Biochemistry 46(25):7599–7606
79. Dai J, Punchihewa C, Ambrus A, Chen D, Jones RA, Yang D (2007) Structure of the intramolecular human telomeric G-quadruplex in potassium solution: a novel adenine triple formation. Nucleic Acids Res 35(7):2440–2450
80. Phan AT, Kuryavyi V, Luu KN, Patel DJ (2007) Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K^+ solution. Nucleic Acids Res 35(19):6517–6525
81. Zhang Z, Dai J, Veliath E, Jones RA, Yang D (2010) Structure of a two-G-tetrad intramolecular G-quadruplex formed by a variant human telomeric sequence in K^+ solution: insights into the interconversion of human telomeric G-quadruplex structures. Nucleic Acids Res 38(3):1009–1021
82. Seenisamy J, Rezler EM, Powell TJ, Tye D, Gokhale V, Joshi CS, Siddiqui-Jain A, Hurley LH (2004) The dynamic character of the G-quadruplex element in the *c-MYC* promoter and modification by TMPyP4. J Am Chem Soc 126(28):8702–8709
83. Phan AT, Modi YS, Patel DJ (2004) Propeller-type parallel-stranded G-quadruplexes in the human c-myc promoter. J Am Chem Soc 126(28):8710–8716
84. Phan AT, Kuryavyi V, Gaw HY, Patel DJ (2005) Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human *MYC* promoter. Nat Chem Biol 1(3): 167–173
85. Mathad RI, Hatzakis E, Dai J, Yang D (2011) *c-MYC* promoter G-quadruplex formed at the 5′-end of NHE III1 element: insights into biological relevance and parallel-stranded G-quadruplex stability. Nucleic Acids Res 39(20):9023–9033
86. Sannohe Y, Sugiyama H (2001) Overview of formation of G-quadruplex structures. In: Current protocols in nucleic acid chemistry. Wiley, New York
87. Xue Y, Z-y K, Wang Q, Yao Y, Liu J, Y-h H, Tan Z (2007) Human telomeric DNA forms parallel-stranded intramolecular G-quadruplex in K^+ solution under molecular crowding condition. J Am Chem Soc 129(36):11185–11191
88. Blume SW, Guarcello V, Zacharias W, Miller DM (1997) Divalent transition metal cations counteract potassium-induced quadruplex assembly of oligo(dG) sequences. Nucleic Acids Res 25(3):617–625

89. Miyoshi D, Nakao A, Sugimoto N (2001) Structural transition of d($G_4T_4G_4$) from antiparallel to parallel G-quartet induced by divalent cations. Nucleic Acids Symp Ser 1(1):259–260
90. Gray RD, Li J, Chaires JB (2009) Energetics and kinetics of a conformational switch in G-quadruplex DNA. J Phys Chem B 113(9):2676–2683
91. Gray RD, Petraccone L, Trent JO, Chaires JB (2009) Characterization of a K^+-induced conformational switch in a human telomeric DNA oligonucleotide using 2-aminopurine fluorescence. Biochemistry 49(1):179–194
92. Miller MC, Le HT, Dean WL, Holt PA, Chaires JB, Trent JO (2011) Polymorphism and resolution of oncogene promoter quadruplex-forming sequences. Org Biomol Chem 9(22): 7633–7637
93. Niermann M, Bolten M, Eimer W (1999) Optimization of the hydrodynamic bead model for the analysis of DNA conformations in solution. J Phys Chem B 103(45):10065–10074
94. Garcia de la Torre J, Navarro S, Lopez Martinez MC, Diaz FG, Lopez Cascales JJ (1994) HYDRO: a computer program for the prediction of hydrodynamic properties of macromolecules. Biophys J 67(2):530–531
95. Bloomfield V, Dalton WO, Van Holde KE (1967) Frictional coefficients of multisubunit structures. I. Theory. Biopolymers 5(2):135–148
96. Petraccone L, Garbett NC, Chaires JB, Trent JO (2010) An integrated molecular dynamics (MD) and experimental study of higher order human telomeric quadruplexes. Biopolymers 93(6):533–548
97. Petraccone L, Spink C, Trent JO, Garbett NC, Mekmaysy CS, Giancola C, Chaires JB (2011) Structure and stability of higher-order human telomeric quadruplexes. J Am Chem Soc 133 (51):20951–20961
98. Petraccone L, Trent JO, Chaires JB (2008) The tail of the telomere. J Am Chem Soc 130(49): 16530–16532
99. García de la Torre J, Huertas ML, Carrasco B (2000) Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. Biophys J 78(2):719–730
100. Ortega A, Amorós D, García de la Torre J (2011) Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. Biophys J 101(4): 892–898
101. García de la Torre J (2001) Hydration from hydrodynamics. General considerations and applications of bead modelling to globular proteins. Biophys Chem 93(2–3):159–170
102. Fernandes MX, Ortega A, López Martínez MC, García de la Torre J (2002) Calculation of hydrodynamic properties of small nucleic acids from their atomic structure. Nucleic Acids Res 30(8):1782–1788
103. Hellman L, Rodgers D, Fried M (2010) Phenomenological partial-specific volumes for G-quadruplex DNAs. Eur Biophys J 39(3):389–396
104. Campbell N, Neidle S (2012) G-quadruplexes and metal ions. Met Ions Life Sci 10:119–134
105. Ida R, Wu G (2008) Direct NMR detection of alkali metal ions bound to G-quadruplex DNA. J Am Chem Soc 130(11):3590–3602
106. Gray RD, Chaires JB (2011) Linkage of cation binding and folding in human telomeric quadruplex DNA. Biophys Chem 159(1):205–209
107. Phan AT, Kuryavyi V, Darnell JC, Serganov A, Majumdar A, Ilin S, Raslin T, Polonskaia A, Chen C, Clain D, Darnell RB, Patel DJ (2011) Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. Nat Struct Mol Biol 18(7):796–804
108. Wei D, Parkinson GN, Reszka AP, Neidle S (2012) Crystal structure of a *c-kit* promoter quadruplex reveals the structural role of metal ions and water molecules in maintaining loop conformation. Nucleic Acids Res 40(10):4691–4700
109. Manning GS (1978) The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. Q Rev Biophys 11(02):179–246
110. Record MT, Anderson CF, Lohman TM (1978) Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity. Q Rev Biophys 11(02):103–178

111. Yphantis DA, Roark DE (1971) Equilibrium centrifugation of nonideal systems. Donnan effect in self-associating systems. Biochemistry 10(17):3241–3249
112. Anantha NV, Azam M, Sheardy RD (1998) Porphyrin binding to quadruplexed T_4G_4. Biochemistry 37(9):2709–2714
113. Haq I, Trent JO, Chowdhry BZ, Jenkins TC (1999) Intercalative G-tetraplex stabilization of telomeric DNA by a cationic porphyrin1. J Am Chem Soc 121(9):1768–1779
114. Freyer MW, Buscaglia R, Kaplan K, Cashman D, Hurley LH, Lewis EA (2007) Biophysical studies of the *c-MYC* NHE III1 promoter: model quadruplex interactions with a cationic porphyrin. Biophys J 92(6):2007–2015
115. Wei C, Jia G, Zhou J, Han G, Li C (2009) Evidence for the binding mode of porphyrins to G-quadruplex DNA. Phys Chem Chem Phys 11(20):4025–4032
116. Fogolari F, Haridas H, Corazza A, Viglino P, Cora D, Caselle M, Esposito G, Xodo L (2009) Molecular models for intrastrand DNA G-quadruplexes. BMC Struct Biol 9(1):64

Top Curr Chem (2013) 330: 211–242
DOI: 10.1007/128_2012_347

Published online: 1 August 2012

Energetics of Ligand Binding to G-Quadruplexes

Concetta Giancola and Bruno Pagano

Abstract G-quadruplex ligands are potential anticancer agents as telomerase inhibitors and potential transcriptional regulators of oncogenes. The search for best-in-class drugs is addressed to identify small molecules able to promote and stabilize G-quadruplex structures. What features should the G-quadruplex ligands possess? They should have selective antiproliferative effects on cancer cells and induce telomerase inhibition or oncogene suppression. One of the main challenges in their design and synthesis is to make the ligands selective for G-quadruplex DNA. These features should be amplified by careful analyses of physico-chemical aspects of G-quadruplex-drug interactions. In particular, the study of the energetics of G-quadruplex-drug interactions can enhance drug design by providing thermodynamic parameters that give quantitative information on the biomolecular interactions important for binding. The main methodologies used to gain information on energetics of binding are based on spectroscopic or calorimetric principles. Spectroscopic techniques such as fluorescence and circular dichroism are rapid and cheap methods, but are not sufficient to characterize completely the thermodynamics of interaction. Calorimetric techniques such as isothermal titration calorimetry offer a direct measure of binding enthalpy, in addition to the stoichiometry and affinity constants. With the complete thermodynamic signature of drug-target interaction, dissecting the enthalpic and entropic components of binding is possible, which can be a useful aid to decision-making during drug optimization.

C. Giancola (✉)
Dipartimento di Scienze Chimiche, Università degli Studi di Napoli "Federico II", Via Cintia, 80126 Napoli, Italy
e-mail: giancola@unina.it

B. Pagano (✉)
Dipartimento di Chimica Farmaceutica e Tossicologica, Università degli Studi di Napoli "Federico II", Via D. Montesano 49, 80131 Napoli, Italy
e-mail: bruno.pagano@unina.it

Keywords Circular dichroism · Fluorescence · G-Quadruplex DNA · G-Quadruplex-drug interaction · Isothermal titration calorimetry · Ligands · Thermodynamics

Contents

1 Introduction: The Importance of Studying the Energetics of Ligand Binding to G-Quadruplexes 212
2 Methodologies to Study the Energetics of G-Quadruplex-Ligand Interactions 213
2.1 ITC: Basic Principles 214
2.2 Spectroscopic Methodologies 219
3 Energetics of End-Stacking or Loop-Interacting G-Quadruplex Ligands 225
3.1 Cationic Porphyrins 225
3.2 The Three Side-Chained Triazatruxene Derivative: Azatrux 231
4 Energetics of Groove Binding G-Quadruplex Ligands 234
4.1 Distamycin and Its Analogs 234
5 Conclusions and Perspectives 238
References 239

1 Introduction: The Importance of Studying the Energetics of Ligand Binding to G-Quadruplexes

Telomeres are nucleoprotein complexes essential in preserving the chromosome ends from catastrophic instability and abnormal chromosome segregation [1]. The telomeric overhang comprises repeats of the DNA motif TTAGGG together with associated telomeric protein (hPOT1), part of a complex of proteins denominated shelterin [2], which regulates telomerase activity in cancer cells. During the normal cellular life, the telomeric G-rich overhangs at the 3′-end are gradually shortened, after each replication cycle, until a critical limit causing the cellular senescence and ultimately apoptosis [3]. These G-rich overhangs are prone to fold back to form G-quadruplex structures, stabilized by consecutive G-tetrads, able to inhibit the enzyme telomerase [4]. Telomerase is a ribonucleoprotein complex with reverse transcriptase activity and its catalytic subunit (hTERT) synthesizes TTAGGG repeats at the end of single-stranded overhangs [5]. The enzyme telomerase is overexpressed in about 85% of cancers, thus favoring telomeres elongation and cell immortalization [6]. Telomerase inhibition with small-molecules or drugs selectively inhibits cancer cell growth and suggests that induction of telomere shortening is a promising anticancer strategy [7]. Telomeric overhang folding into G-quadruplex structure inhibits telomerase activity by subtracting the substrate of the enzyme, essential for hybridization with RNA template in the first step of catalytic cycle [4].

For these reasons, many researchers have focused their studies on the search for small molecules able to promote and stabilize G-quadruplex structures, so acting as telomerase inhibitors.

In a review on the current status of telomeric G-quadruplexes as therapeutic targets in human cancer, Neidle draws attention to the evidence that researchers need to have in identifying G-quadruplex ligands as drugs [8]. Among these, high-affinity G-quadruplex binding, with K_b value at least 10^6 M^{-1}, is essential. The specificity for quadruplex binding over duplex DNA is also critical, with an association constant of at least 10^2 greater for G-quadruplex over duplex being necessary. Physico-chemical assay are pivotal for such in vitro evidence. Biological assays are also very important to gain evidence about selective inhibition of cancer cell lines, senescence, competitive inhibition of hPOT1 binding in cell, and telomeric uncapping in cell from hTERT.

In this chapter, the discussion will be focused on physico-chemical assays devoted to study the energetic aspects of the binding of small molecules to G-quadruplex structures. The study of energetics of G-quadruplex–drug interactions can be useful to address the drug design and, specifically, thermodynamic parameters can give quantitative information on biomolecular interactions providing the "affinity" of the interactions under equilibrium conditions.

2 Methodologies to Study the Energetics of G-Quadruplex-Ligand Interactions

A number of biophysical methodologies are available to identify new natural or synthetic ligands as putative drugs and to investigate their interaction with G-quadruplexes. These methodologies can be used simply for evaluating properties like ligand affinity or to obtain more detailed information about thermodynamics and stoichiometry of the interactions. Due to the high degree of structural polymorphism in G-quadruplexes, it is essential to determine the binding sites of the ligand and its mode of interaction. The challenge in G-quadruplex binder design and synthesis is to make the ligand selective for G-quadruplex structures in comparison to duplex DNA. The biophysical methods being used to investigate these interactions should be able to detect and measure the ligand selectivity for G-quadruplexes over duplexes.

The main methodologies devoted to gaining information on energetics of binding are based on spectroscopic or calorimetric principles. The spectroscopic techniques such as fluorescence and circular dichroism (CD) are rapid methods and require small amounts of material. A calorimetric technique as isothermal titration calorimetry (ITC) allows a deeper insight into the energetics aspects of the studied interactions. The major difference between these methodologies is that in ITC the binding enthalpy is directly measured, whereas in CD and fluorescence experiments the binding enthalpy is calculated by van't Hoff analysis. Furthermore, ITC allows

dissecting the enthalpic and entropic contributions associated with the binding of a compound to a target, thus defining its thermodynamic signature [9, 10]. In principle, high affinity binding requires favorable contributions from both enthalpy and entropy, but these two contributions often work in opposition because enthalpic optimization can counterbalance a loss in the entropic contribution. The optimization of the entropic contribution to the binding energy of a ligand to its target is generally much easier than refining the enthalpic contribution, since increasing the drug lipophilicity could increase the entropic contribution [11].

In the following paragraphs, a description of the basic principles of the aforementioned biophysical methodologies is reported.

2.1 ITC: Basic Principles

ITC is a valuable tool for characterizing the energetic aspects of interactions between G-quadruplexes and other biomolecules, including small ligands, due to its general applicability and precision. A number of reviews devoted to this methodology describe the general principles of the ITC technique and its application to the study of the energetics of bio-macromolecules interactions [12–14]. Thanks to the great current interest in developing molecules that can selectively bind to G-quadruplexes, the number of ITC studies on the interaction between small ligands and G-quadruplex structures is rapidly growing [15–18]. Many small molecules that bind to G-quadruplexes have proven to be effective therapeutic agents, although the exact mode of binding and the nature of thermodynamic forces that regulates G-quadruplex-ligand interactions are often poorly understood. Drug development of lead compounds can be greatly enhanced by detailed knowledge of the thermodynamics of binding to the target [19]. Since ITC has the ability to determine the different energetic contributions to the binding affinity, it can provide a unique bridge between computational and experimental analysis.

ITC is a high-accuracy method for measuring binding affinity and stoichiometry; moreover, it directly measures the heat of interaction (enthalpy change, $\Delta_b H^\circ$), at a constant temperature, on titrating two substances of known concentration forming an equilibrium complex. It is the only methodology capable of directly quantifying both enthalpic and entropic components of the interaction. A typical ITC experiment is carried out by the stepwise titration of one component of the complex (for example, a ligand) in the sample cell containing the other component (for example, a G-quadruplex molecule). The binding events involve the release or absorption of heat (the enthalpy change, $\Delta_b H^\circ$) and the analysis of these thermal effects allows a full thermodynamic characterization of the reaction. At any point in the titration the amount of free or bound ligand can be determined, establishing the equilibrium binding constant (K_b). ITC experiments cover an optimum range of binding constants in the range 10^4–10^8 M^{-1} corresponding to dissociation constants ($K_b = 1/K_d$) of 100 μM to 10 nM. The temperature is kept constant throughout the entire ITC experiment; consequently the free energy of the binding reaction

($\Delta_b G^\circ$) can be determined from $\Delta_b G^\circ = -R\,T \ln K_b$, where R is the gas constant. The change in entropy ($\Delta_b S^\circ$) can be calculated as $\Delta_b S^\circ = (\Delta_b H^\circ - \Delta_b G^\circ)/T$. The enthalpic and entropic contributions to the Gibbs energy of the binding process are related to the structural parameters and can be used as a guide to drug design [9]. In addition, those contributions can be used to confirm the structure-based computational predictions of binding energetics and to find structure-energy correlations. Among the advantages of ITC is that it allows the accurate determination of thermodynamic parameters to be made with no requirement for chemical modification, labeling, or immobilization of the G-quadruplex and/or ligands.

Recent advances in instrumentation, together with a broad availability of automated and high-sensitivity ITC instruments, have enhanced the opportunity to provide additional information to drug discovery and optimization processes. The instruments of latest generation, the nano-ITCs, can detect heat effects as small as 0.4 mJ, allowing titrations to be carried out with nanomoles of substance.

The most widely used calorimeters feature a reference cell, containing water or buffer solution, and a sample cell, containing one of the interacting molecules. The two cells are constantly kept in thermal equilibrium with the bath during the experiment. Thermoelectric device sensors that surround the vessel precisely measure the heat flow between the sample cell and the isothermal block. The heat flow is monitored as a function of time. The nano-ITC uses a differential power compensation design for maximum sensitivity and responsiveness. In a typical experiment, a ligand contained in a precision syringe (25–250 μL) is added by several computer-controlled injections (1–20 μL) into the sample cell, resulting in several peaks in the thermogram (Fig. 1a). At the end of the titration, ligand concentration should be in excess with respect to G-quadruplex binding sites. The ligand and macromolecule should be in the same buffer and salt conditions to avoid any large background heat effect arising from dilution and mixing of non-identical buffers. Further, when conducting ITC experiments with nucleic acids, high sample purity is required. It is very important that the sample is better than 99% pure. Impurities, especially similar sequences, could compete with the target sequence, thus giving a wrong stoichiometry and/or a more complex binding curve than a simple sigmoid.

2.1.1 ITC: Thermodynamic Analysis

The analysis of extremely small thermal effects arising from the binding allows a full thermodynamic characterization of the interaction and provides fundamental information about the molecular interactions driving the process.

In ITC, heat is either generated or taken in (exo- or endothermic process) upon making an injection of ligand. The heat effect is detected by semiconductor thermopiles and the instrument applies thermal power (microjoules per second) to compensate for the ligand induced heat effect and thus to bring the cells into thermal equilibrium. Figure 1a illustrates the raw data produced by the sequence of injections as a series of peaks measured as power (microjoules per second) vs

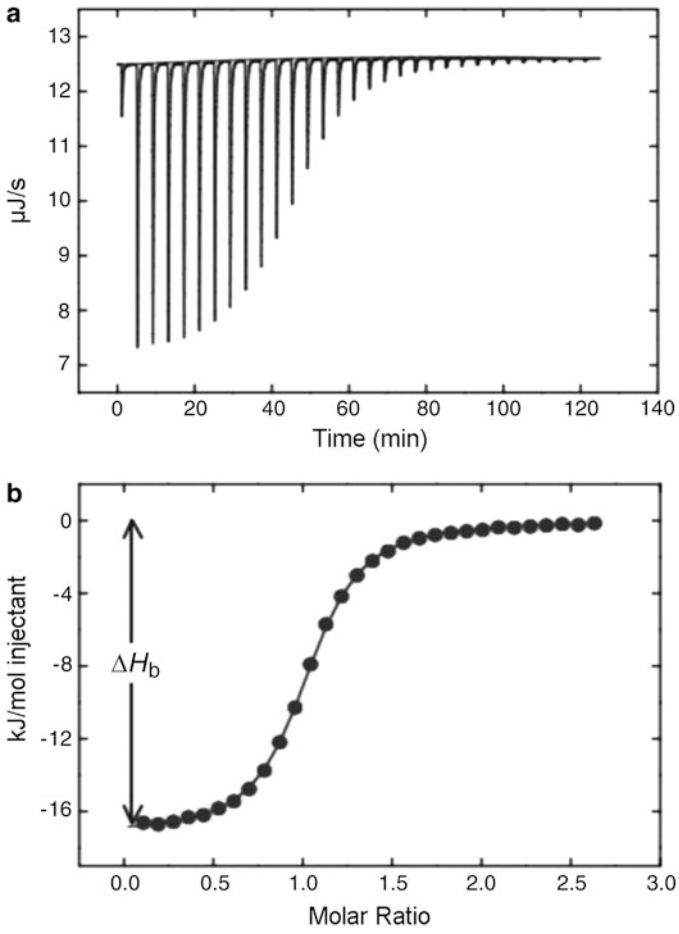


Fig. 1 (**a**) Typical raw data produced by the sequence of injections for an ITC experiment. (**b**) Binding isotherm obtained after integration of each injection peak

time. The magnitude of the heat per injection decreases because initially almost all injected ligand binds to the target molecule but, after each successive injection, the amount of uncomplexed G-quadruplex available progressively decreases. When complete saturation is achieved, the magnitude of the peaks becomes progressively smaller and similar. Such heat corresponds to the ligand dilution heat. Before thermodynamic analysis the heat of dilution must be subtracted. It could be measured in separate experiments where the ligand is injected into the buffer alone. A typical plot of the integrated peaks as a function of the ligand/G-quadruplex molar ratio generates a curve with a sigmoidal shape that facilitates the estimation of the midpoint of the reaction process, and thus the stoichiometry of the binding reaction (Fig. 1b). The resulting binding isotherm can then be obtained

by using the appropriate binding site model to give the equilibrium binding constant (K_b) and the stoichiometry.

For a single site system, where the complex QuL is in equilibrium with the G-quadruplex (Qu) and the ligand (L), the binding process is described by

$$\mathrm{Qu} + L \leftrightarrow \mathrm{Qu}L$$

and the equilibrium constant K_b can be described by the equation

$$K_b = \frac{[\mathrm{Qu}L]}{[\mathrm{Qu}]\,[L]}. \tag{1}$$

The chemical reaction generated by each injection either releases or absorbs a certain amount of heat (q) proportional to the amount of ligand that binds to the macromolecule in a particular injection ($V \times [\Delta L]_B$) and the characteristic binding enthalpy (ΔH) for the reaction

$$q = \Delta H\, \Delta[L]_B\, V, \tag{2}$$

where $\Delta[L]_B$ is the increase in the concentration of bound ligand after the ith injection and V the volume of the sample cell [20].

The total heat (Q) of the binding reaction is

$$Q = V\, \Delta H \sum \Delta[L]_B = V\, \Delta H\, [L]_B. \tag{3}$$

$[L]_B$ can be written as function of free ligand concentration $[L]$ and total G-quadruplex concentration $[\mathrm{Qu}]_T$ through K_b and the mass conservation relationships

$$[\mathrm{Qu}L] = K_b[\mathrm{Qu}]\,[L] \tag{4}$$

and

$$[L]_T = [\mathrm{Qu}L] + [L] = K_b\,[\mathrm{Qu}]\,[L] + [L] = [L]\,(K_b[\mathrm{Qu}] + 1), \tag{5}$$

$$[\mathrm{Qu}]_T = [\mathrm{Qu}L] + [\mathrm{Qu}] = K_b[\mathrm{Qu}][L] + [\mathrm{Qu}] = [\mathrm{Qu}](K_b[L] + 1), \tag{6}$$

where $[L]_T$ is the total ligand concentration.

The concentration of free Qu can be obtained from (6):

$$[\mathrm{Qu}] = \frac{[\mathrm{Qu}]_T}{K_b[L] + 1} \tag{7}$$

that substituted in (5) gives

$$[L]_B = [QuL] = [Qu]_T - [Qu] = [L]_T - [L] = [L]\left(\frac{K_b[Qu]_T}{K_b[L]+1}+1\right) - [L]$$
$$= \frac{K_b[Qu]_T[L]}{K_b[L]+1}. \quad (8)$$

Equation (8) can be introduced into (3) and, for a more general case of n_i independent and equivalent sites:

$$Q = V\sum_i \Delta H_i\,[L]_{B,i} = V\,[Qu]_T \sum_i \left(\frac{n_i \Delta H_i K_{b,i}[L]}{1+K_{b,i}[L]}\right). \quad (9)$$

The total ligand concentration is known, whereas the concentration of free ligand is unknown. For this reason it is necessary to write an analytical expression of the free ligand concentration that contains the total ligand concentration $[L]$ and the total G-quadruplex concentration $[Qu]_T$.

From the mass conservation relation of ligand and with the relation obtained for $[L]_B$ the following equation can be obtained:

$$[L]_T = [L] + [L]_B = [L] + \left(\frac{K_b[Qu]_T[L]}{K_b[L]+1}\right) = \frac{K_b[L]^2 + [L] + K_b[Qu]_T[L]}{K_b[L]+1}. \quad (10)$$

If one multiplies the first and the last term by $K_b[L] + 1$, obtains the following second degree equation:

$$[L]_T K_b[L] + [L]_T = K_b[L]^2 + [L] + K_b[Qu]_T[L]. \quad (11)$$

The only acceptable solution that satisfies this equation is the following:

$$[L] = \frac{-(1+[Qu]_T K_b - [L]_T K_b) + \sqrt{(1+[Qu]_T K_b - [L]_T K_b)^2 + 4[L]_T K_b}}{2K_b}. \quad (12)$$

Finally, by introducing the last equation in (9), one can obtain the relationship between the heat and the total macromolecule and ligand concentrations at each injection. A value for K_b is initially estimated and then the concentration of bound complex is calculated for each injection. In combination with the measured heat, these values are used to determine the average of ΔH. The ΔH and the calculated concentration are then used to determine an expected heat per injection, and the error square sum between the measured and expected heat for each peak is calculated. The value of K_b is then adjusted and the process repeated until a minimum error square sum is obtained. However, in some cases a more complex approach to data analysis is required if there are multiple sites with differing

binding constants. A more extensive description of ITC data analysis can be found in [21] and references therein.

2.2 Spectroscopic Methodologies

2.2.1 Fluorescence Spectroscopy

Fluorescence spectroscopy is a biophysical methodology particularly useful for studying G-quadruplex-ligand interactions. Binding constant and stoichiometry can be easily obtained with small amounts of DNA and ligands (micromolar or less). Fluorescence is based on the emission of a photon upon relaxation from an electronically excited singlet state to the ground state subsequent to the absorption of a photon by the fluorescent molecule (fluorophore). The sample is excited by irradiation at a certain wavelength and emits radiation at a different wavelength. The process takes place on a much slower time scale (~nanoseconds) than absorption (~femtoseconds) allowing a much wider range of interactions to influence the fluorescence emission spectrum. The emission spectrum provides information for both qualitative and quantitative analysis.

Fluorescence spectrometers detect the relative fluorescence, specifically the emission intensity of the system, which is a function of the instrument (intensity of the excitation source, sensitivity of the detection system, etc.) and of the measurement conditions (fluorophore concentration, temperature, and solution conditions) and it is often given in arbitrary units. The relative fluorescence is most accurately obtained from integration over the wavelengths of the emission spectrum but, for many purposes, it is sufficient to determine the fluorescence intensity at the maximum of the fluorescence spectrum. In some cases, the fluorescence quantum yield is detected, which is the ratio of photons absorbed to photons emitted through fluorescence. For an extensive treatment of fluorescence spectroscopy see [22].

The intrinsic fluorescence of most G-quadruplex-forming oligonucleotides is too low to be detected, although in recent papers it has been established that G-quadruplex structures can be characterized and distinguished by their intrinsic fluorescence spectra [23–25]. It has been found that G-quadruplexes have higher intrinsic fluorescence quantum yields than their denatured oligonucleotides [25]. Further, fluorescence and fluorescence excitation spectra were acquired to follow the formation of G-quadruplexes as a function of cations and temperature. Indeed, the temperature-dependent fluorescence spectra of different G-quadruplexes exhibit characteristic patterns.

However, to enhance the fluorescence of G-quadruplex, a suitable fluorescent reporter group is often incorporated to impart fluorescence to the polynucleotide structure. Highly fluorescent adenine analog, 2-aminopurine (2-AP), has been widely used as a probe for loop structures in G-quadruplex DNA. Its quantum yield is highly sensitive to environmental conditions and it can be used to probe the interaction of G-quadruplexes with small organic molecules. The 2-aminopurine can be incorporated in any G-quadruplex nucleic acid without changing the

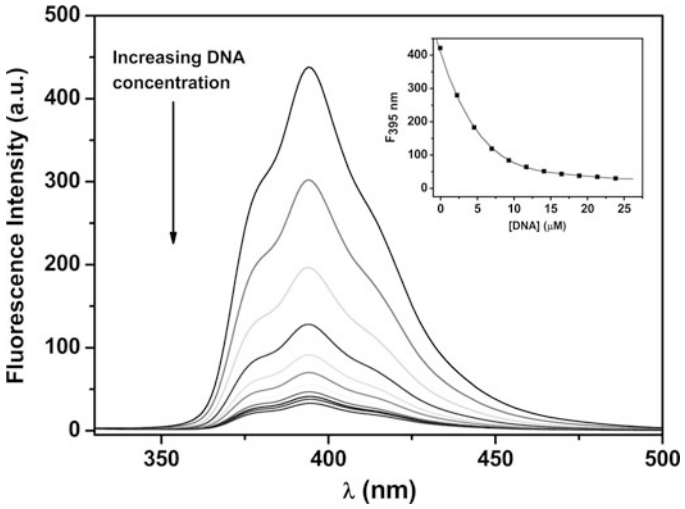


Fig. 2 Example of fluorescence titration experiment. Ligand fluorescence spectra on increasing G-quadruplex concentration are shown. The intensity of fluorescence decreases on increasing G-quadruplex concentration. *Inset*: binding curve obtained by plotting the fluorescence intensity at the maximum of the emission spectrum as a function of DNA concentration

G-quadruplex conformation and only slightly affecting its stability. The 2-AP fluorescence intensity tends to increase with solvent exposure and decrease with base stacking. The overall fluorescence of a G-quadruplex changes if a small molecule interacts in the proximity of 2-AP.

On the other hand, a number of reported G-quadruplex ligands have intrinsic fluorescence properties and show a large change in their fluorescence intensity upon DNA binding [18, 26, 27]. In these cases, fluorescence titration experiments can be performed following the variation of the ligand fluorescence on increasing G-quadruplex concentration (Fig. 2). The binding curve obtained by plotting the fluorescence intensity at the maximum of the ligand emission spectrum as function of the G-quadruplex concentration can be fitted to a specific binding model to determine the binding stoichiometry and affinity constant [27, 28].

An alternative methodology based on fluorescence is the FID (fluorescent intercalator displacement) assay, involving the displacement of an "on/off" fluorescence probe, Thiazole Orange (TO), from G-quadruplex or duplex DNA receptor by increasing amounts of a putative ligand that competes for the TO binding site (Fig. 3) [29]. TO is not fluorescent when free in solution, but it is strongly fluorescent when bound to DNA. Its binding constant for the d[AGGG(TTAGGG)$_3$] G-quadruplex-forming oligonucleotide is $3 \times 10^6\ M^{-1}$ [30]. The G-quadruplex affinity of a candidate compound can be evaluated through its ability to displace TO from G-quadruplex, but thermodynamic parameters cannot be determined. However, selectivity measurements can be accomplished by comparing the ability of the ligand to displace TO from G-quadruplex and duplex structures.

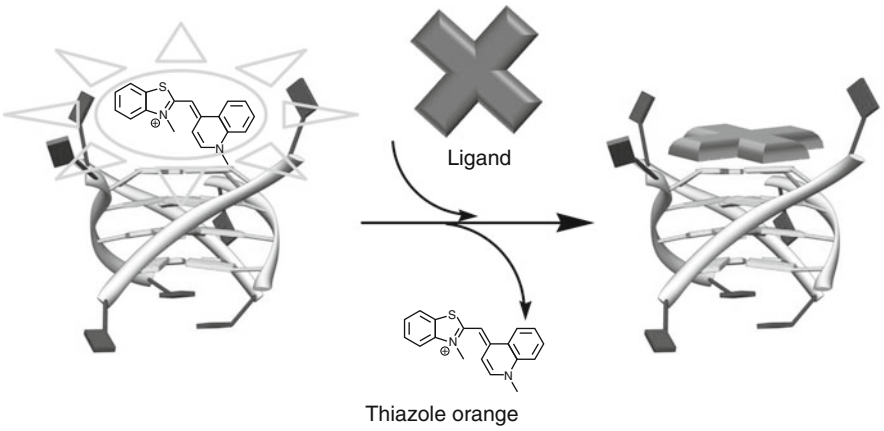


Fig. 3 Schematic representation of the FID assay. The thiazole orange, which is strongly fluorescent when bound to G-quadruplex, becomes not fluorescent when displaced from the DNA by a ligand with higher G-quadruplex affinity

This method does not require fluorescence of ligands, employs readily available materials and equipment, and can be easily implemented with 96-well plates reader for rapid high throughput screening ligands [29]. A limitation of this method may arise from fluorescent ligands, whose fluorescence could interfere with TO excitation and emission maxima. Another limitation is that some ligands could bind G-quadruplex at alternate sites without displacing TO, leading to underestimation of the ligand affinity.

A valid methodology for screening putative G-quadruplex-ligands is based on the fluorescence resonance energy transfer (FRET) effect. A FRET experiment involves a G-quadruplex-forming oligonucleotide covalently linked to two fluorophore probes, a donor, and an acceptor, with the request that the emission spectrum of the donor probe overlaps the absorption spectrum of the acceptor probe. If it does, the donor fluorophore transfers its excitation energy to the acceptor fluorophore, in a non-radiative manner, with an efficiency depending on the spectral properties, distance, and relative orientations. The distance within which the non-radiative energy transfer takes place is known as the "Förster distance" (R_0), and it is characteristic for the donor–acceptor pair. It is usually within the range 10–80 Å [31].

Mergny et al. developed a FRET-based melting assay: the ligand affinity for a G-quadruplex is evaluated by measuring the increase in the melting temperature (T_m) of the G-quadruplex induced by the presence of the ligand [32, 33]. In the "FRET melting assay" the increase in temperature leads to G-quadruplex denaturation (melting). Melting causes the ligand displacement on the G-quadruplex folded–unfolded equilibrium around the melting temperature and an increasing distance between the probes leads to an increase in fluorescent energy. Usually, to monitor the thermal G-quadruplex unfolding, fluorescein (FAM) can be used as a fluorescent probe and dabcyl as a quencher. Alternatively, an acceptor for FRET, e.g., tetramethylrhodamine (TAMRA), can be used. These groups are attached to

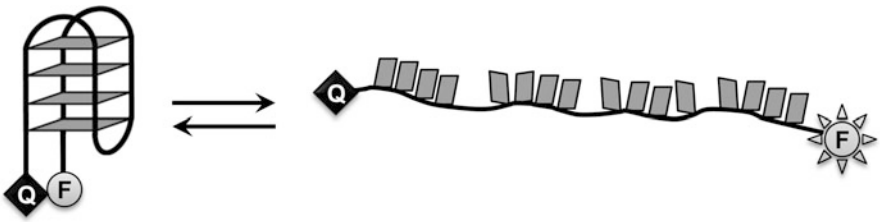


Fig. 4 Schematic representation of the FRET-based assay. Intramolecular folding of an oligonucleotide into a G-quadruplex leads to fluorescence resonance energy transfer between a fluorescent probe and a quencher covalently attached to the 5′ and 3′ ends of the DNA, respectively. The fluorescence intensity of the fluorescent probe depends on its distance from the quencher. This distance changes considerably in the G-quadruplex folding/unfolding process

the 5′-end and 3′-end of the oligonucleotide, respectively, as shown in Fig. 4. The fluorescence intensity of the fluorescent probe depends on its distance from the quencher (or from the FRET acceptor); this distance significantly changes in the G-quadruplex folding/unfolding process. The large difference in the fluorescence properties of the folded and unfolded G-quadruplex allows one to obtain well resolved melting curves. FRET melting assays can also detect ligand selectivity. It can be measured by adding a non-tagged (non-fluorescent) duplex DNA into the test solution containing ligand bound to G-quadruplex-forming oligonucleotide and then conducting the melting experiments [34]. FRET is a rapid and convenient method for high-throughput screening. On the other hand, the method can generate false positives due to quenching of the donor emission by the tested ligand, or false negatives because the G-quadruplex folds with the ligand in a configuration where the distance prevents quenching.

In conclusion, FID and FRET are useful methodologies for a preliminary screening to select putative ligands that selectively bind to G-quadruplex, but cannot be exploited to extract intrinsic binding properties such as stoichiometry and affinity constant or thermodynamic parameters. To achieve those parameters, steady-state fluorescence should be more conveniently utilized.

2.2.2 Circular Dichroism Spectroscopy

Circular dichroism (CD) is the phenomenon detected when an optical active sample interacts with plane-polarized light composed of right-circularly and left-circularly polarized components. Specifically, CD spectroscopy measures the difference in absorbance of right-circularly and left-circularly polarized light after they go through the sample [35]. Since the two components have different amplitude, the light in output results elliptically polarized. Recording the CD at different wavelength allows one to obtain the CD spectrum, which is strongly indicative of the secondary structure content of macromolecules and, among these, of the G-quadruplexes. For its sensitivity to stereochemical variations, CD has emerged as an important technique for studying subtle conformational changes.

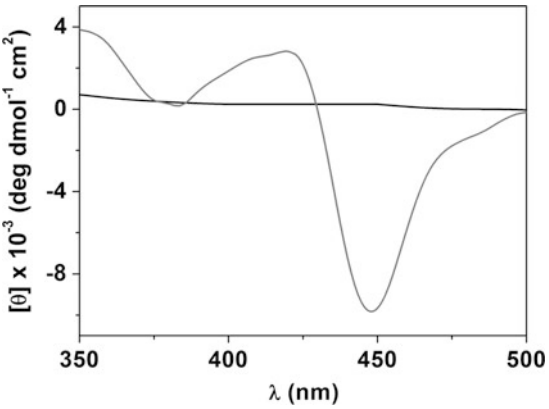


Fig. 5 Example of induced CD signal (*gray line*) of a ligand bound to a DNA G-quadruplex. In absence of the G-quadruplex (*black line*) the ligand does not show CD signal

CD is therefore the most useful technique to compare G-quadruplex conformations and to detect changes when the solvent or temperature is changed. It is extremely useful in studies of G-quadruplex conformations, since the CD features are mainly influenced by the type of stacking interactions between adjacent G-quartets [36, 37].

For this reason, CD can be used to discriminate among different folding topologies of G-quadruplexes [36, 38, 39]. Although the folding topology of G-quadruplex strands is very complex and many types of G-quadruplexes have been reported, there are two basic types of CD spectra that are characterized by distinctive marker bands: the spectra of a parallel structure (all strands with the same 5′ to 3′ orientation) show a positive band at ca. 260 nm and a negative peak at ca. 240 nm [40, 41], whereas the spectra of an antiparallel G-quadruplex have a negative band at ca. 260 nm and a positive band at ca. 290 nm [40, 42].

A more extensive description of CD applied to G-quadruplex structures can be found in [86]. In the present chapter, CD is regarded as methodology to follow ligand binding to G-quadruplexes and to evaluate, in some cases, the selectivity of the ligand for a specific G-quadruplex topology. The titration of G-quadruplexes with ligands could induce changes in the CD spectrum of a G-quadruplex. Such changes could be correlated to the properties of the ligand under investigation. In a typical CD titration, the ligand is added stepwise into the cell containing the G-quadruplex sample, and the CD spectrum is collected after each addition. The gradual appearance of characteristic bands in particular regions of the CD spectrum provides specific information concerning the binding site and the binding mode of the ligand. A typical titration curve can be obtained detecting the molar ellipticity of a maximum or minimum in the spectra against ligand concentration or ligand/quadruplex ratio.

In addition, further information on G-quadruplex/ligand interactions is provided by the induced circular dichroism (ICD). ICD signal is detected when a CD-inactive non-chiral ligand binds to a chiral host, such as DNA G-quadruplex, acquiring, in the bound state, a CD signal due to the chiral environment [43] (Fig. 5).

As far as we know, ligands generally interact with G-quadruplex structures through groove and end-stacking binding modes. It was shown that the sign of the ICD spectra of molecules bound to nucleic acids provides important information about their way of interaction: a positive ICD band can be indicative of groove binding, whereas a negative ICD signal (or a slightly intense positive band) can be indicative of end-stacking interaction [44, 45].

2.2.3 Spectroscopic Binding Curve: Thermodynamic Analysis

The binding curve resulting from G-quadruplex-ligand interaction can be obtained by plotting a spectroscopic signal as function of G-quadruplex or ligand concentration or as function of ligand/G-quadruplex concentration ratio. In UV and fluorescence spectroscopies the experimental variables are the absorbance and fluorescence emission intensity at the maximum of titration spectra, respectively. In the CD spectroscopy, molar ellipticity or ICD signal of a maximum or minimum in the titration spectra can be followed as a function of ligand concentration.

In any case, the experimental variable can be considered as the sum of the weighted contributions from free and bound ligand:

$$Y = (1 - \alpha_b)Y_0 + \alpha_b Y_b, \tag{13}$$

where Y is the value of the experimental variable at each titrant concentration, Y_0 and Y_b are the values of the variable at the initial and final states of titration, and α_b is the mole fraction of ligand in bound form. Assuming 1:1 stoichiometry for the interaction and no intermediate states, it can be shown that

$$[L]_0\alpha_b^2 - ([L]_0 + [\mathrm{Qu}] + 1/K_b)\alpha_b + [\mathrm{Qu}] = 0, \tag{14}$$

where K_b is the binding constant, $[L]_0$ is the total ligand concentration, and [Qu] is the added DNA concentration. From combination of previous equation and considering

$$\alpha_b = \frac{Y - Y_0}{Y_b - Y_0}, \tag{15}$$

it can be shown that:

$$\Delta Y = \left(\frac{\Delta Y_{max}}{2\,[L]_0}\right)\left\{([L]_0 + [\mathrm{Qu}] + 1/K_b) - \sqrt{([L]_0 + [\mathrm{Qu}] + 1/K_b)^2 - 4[L]_0[\mathrm{Qu}]}\right\}, \tag{16}$$

where $\Delta Y = Y - Y_0$ and $\Delta Y_{max} = Y_{max} - Y_0$. Y_{max} is the maximum value of the experimental variable [46]. In the general case of a stoichiometry n:1, where n is the

number of equivalent and independent sites on G-quadruplex structure, the following equation can be utilized [27, 46, 47]:

$$\alpha = (1/2[L]_0)\left\{ \left([L]_0 + n[\mathrm{Qu}]_0 + 1/K_\mathrm{b}\right) - \sqrt{\left([L]_0 + n[\mathrm{Qu}]_0 + 1/K_\mathrm{b}\right)^2 - 4[L]_0 n[\mathrm{Qu}]_0} \right\}. \tag{17}$$

3 Energetics of End-Stacking or Loop-Interacting G-Quadruplex Ligands

When small molecules bind to duplex DNA there are two predominant binding modes – intercalation and groove binding. From calorimetric measurements it is possible establish enthalpic and entropic contributions to the binding. Generally, the binding of an intercalator to duplex is driven by a large favorable enthalpy change, with an unfavorable entropy decrease. Conversely, the binding of a groove binder is characterized by a large favorable increase of entropy and a small (favorable or unfavorable) enthalpy change [48]. For G-quadruplex structures, groove binding, end-stacking, or loop interactions were found to be the predominant binding modes, whereas intercalative binding of the drug between two adjacent G-tetrads was only hypothesized in a few cases. Here, some examples of drugs that interact with those different binding modes are considered.

3.1 Cationic Porphyrins

Cationic porphyrins are largely investigated as G-quadruplex-interactive drugs, especially by physico-chemical methodologies. These compounds are known to bind to and stabilize different types of G-quadruplexes and, in some cases, to facilitate G-quadruplex formation [49]. The cationic *meso*-tetrakis-(*N*-methyl-4-pyridyl)-porphyrin (TMPyP4, Fig. 6a) has been extensively investigated, since it induces telomerase inhibition upon binding to telomeric DNA G-quadruplexes [50]. Furthermore, it has been shown to downregulate the expression of the *c-myc* oncogene by stabilizing the G-quadruplex structure adopted by the guanine-rich segment of the P1 gene promoter [51].

The geometric features of the TMPyP4 ring have been assumed to be appropriate to interact with the G-tetrads of G-quadruplex structures through π–π stacking interactions. However, the binding modes as well as the binding stoichiometry in TMPyP4-G-quadruplex complexes continue to be controversial.

Several different models of TMPyP4 binding to G-quadruplexes have been proposed on the basis of structural, spectroscopic, calorimetric, or computational

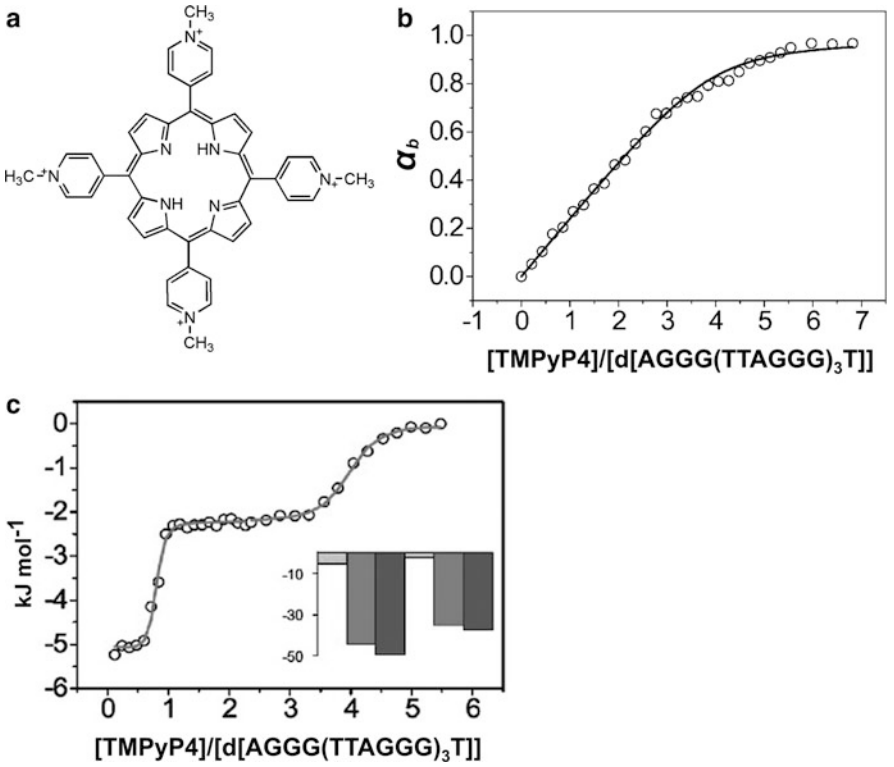


Fig. 6 (**a**) Chemical structure of TMPyP4. (**b**) Example of CD titration experiment on the interaction between TMPyP4 and a G-quadruplex. The *open circles* represent the experimental data obtained following the change of CD signal at 290 nm, upon TMPyP4 addition. The *line* represents the theoretical curve obtained on the basis of the appropriate model. (**c**) Example of ITC binding isotherm for titration of a G-quadruplex with TMPyP4. The *open circles* represent the experimental data obtained by integrating the raw data and subtracting the heat of ligand dilution into the buffer. The *line* represents the best fit obtained with the appropriate model. *Inset* shows the thermodynamic signature of the corresponding interaction. The enthalpic contribution (ΔH°) is shown in *light gray*, the entropic one ($-T\Delta S^{\circ}$) in *gray*, and the Gibbs energy change (ΔG°_{298K}) in *dark gray*

studies. Proposed binding modes include intercalative binding of the drug between two adjacent G-tetrads and/or stacking with the external G-tetrads or with the loops of the G-quadruplexes [52–60]. The clarification of binding mode is further complicated by the intrinsic structural polymorphism of G-quadruplex sequences.

Further, there is a debate on the binding stoichiometry of TMPyP4 to G-quadruplex structures and, especially for the G-quadruplexes from human telomeric sequences, the issues have not been fully resolved. Many groups studied TMPyP4-G-quadruplex interactions under different conditions and by different methodologies. More than a decade ago, Wheelhouse et al. determined by UV absorption spectroscopy a stoichiometric ratio of 2:1 for the binding of TMPyP4

to the 22-mer d[AGGG(TTAGGG)$_3$] G-quadruplex in 100 mM Na^+ buffer [50], whereas, more recently, Wei et al., using the same technique, determined a 4:1 ratio for the same interaction in identical solution conditions [57]. Furthermore, Haq et al. determined, by ITC, the stoichiometry and thermodynamics for binding of TMPyP4 to the d[AGGG(TTAGGG)$_3$] G-quadruplex structure in two different buffered solutions containing either K^+ or Na^+ [52]. They found that the binding stoichiometry was the same for the two solutions, in agreement with the 2:1 stoichiometry value previously determined for the human sequence by Wheelhouse et al. On the other hand, the interaction with d[AGGG(TTAGGG)$_3$] in K^+ was characterized by a single binding event, whereas the binding in Na^+ solution was characterized by an initial binding event requiring only low [TMPyP4]/[DNA] ratios to reach saturation, followed by a secondary process requiring a much higher ligand concentration for saturation. Their study revealed that the interactions can be affected by the ionic environment, which may be one of the key factors determining the binding mode of TMPyP4 and the stoichiometry of the complexes, as also observed for other ligands [61]. The binding of TMPyP4 to the human telomeric G-quadruplex formed by the longer sequence d[TAGGG(TTAGGG)$_3$T] was characterized by Zhang et al. for a range of K^+ concentrations [49]. They found that TMPyP4 forms complexes with a [drug]/[G-quadruplex] ratio of 5:1, 5:1, and 3:1 in the presence of 0, 10, and 100 mM K^+, respectively. They also found that an increase in K^+ concentration decreased the binding affinity of TMPyP4 and induced hybrid-type G-quadruplex transition into an anti-parallel structure. A conformational transition induced by TMPyP4 was also suggested by Gaynutdinov et al. using ^{125}I-radioprobing for the d[GGG(TTAGGG)$_3$^{125}I-CT] DNA sequence labeled with ^{125}I [59]. They showed that d[GGG(TTAGGG)$_3$^{125}I-CT] was a mixture of two or more conformations in K^+ solution, but TMPyP4 stabilized a basket conformation. Further, they also showed that, in the presence of Na^+, the same sequence adopted a basket conformation, which did not change upon drug addition. This disagrees with a more recent study by Gray et al. which showed that TMPyP4 addition promoted the conversion of d[AGGG(TTAGGG)$_3$] from the basket to the hybrid form [62].

Arora and Maiti investigated the energetics of TMPyP4 binding to three G-quadruplex structures differing in loop orientation [60]. The G-quadruplexes used in their study were the 21-mer d[$G_3(T_2AG_3)_3$] from human telomeric sequence, the 22-mer d($G_4AG_3TG_4AG_3TG_4$) from the promoter region of *c-myc*, and the 21-mer d($G_3AG_3CGCTG_3AG_2AG_3$) from the promoter region of *c-kit*. The telomeric G-quadruplex adopted antiparallel conformation, whereas the *c-myc* and *c-kit* G-quadruplex structures adopted an intramolecular parallel-stranded conformation. ITC experiments showed biphasic profiles consistent with two independent binding processes for all the G-quadruplexes. The stoichiometry of interaction for the first and second binding was found to be 1:1 and 2:1 (ligand/G-quadruplex), respectively (Table 1). The first binding process was stronger than the second one and the binding processes were enthalpically driven in all cases. On the basis of the overall energetic picture, they hypothesized that the stronger binding was due to an end-stacking mode favored by the interaction of aromatic planar rings of TMPyP4

Table 1 Thermodynamic parameters for cationic porphyrins binding to G-quadruplexes obtained by ITC at 25 °C

		First binding event					Second binding event				
G-Quadruplex	Buffer	n_1[a]	K_b (M^{-1})	ΔH° (kJ mol^{-1})	$-T\Delta S^\circ$ (kJ mol^{-1})	ΔG° (kJ mol^{-1})	n_2[a]	K_b (M^{-1})	ΔH° (kJ mol^{-1})	$-T\Delta S^\circ$ (kJ mol^{-1})	ΔG° (kJ mol^{-1})
$d(G_2T_2G_2TGTG_2T_2G_2)$[b]	K^+	1	1.8×10^5	−40.2	10.0	−30.1	–	–	–	–	–
$d[AG_3(T_2AG_3)_3]$[b]	K^+	2	2.8×10^4	−17.6	−8.0	−25.5	–	–	–	–	–
$d[AG_3(T_2AG_3)_3]$[b]	Na^+	1	3.3×10^4	−28.4	2.5	−25.9	1	~0.2×10^4	~−25	n.d.	n.d.
$[d(T_4G_4)]_4$[b]	K^+	3	7.7×10^4	−38.1	10.0	−28.0	–	–	–	–	–
$[d(T_4G_4)]_4$[b]	Na^+	1	1.6×10^6	−28.0	−7.5	−35.6	2	4.4×10^4	−105.8	79.5	−26.3
$d[G_3(T_2AG_3)_3]$[c]	K^+	1	4.0×10^6	−7.9	−29.7	−37.6	2	0.5×10^6	−17.6	−15.0	−32.6
$d(G_4AG_3TG_4AG_3TG_4)$[c]	K^+	1	7.0×10^7	−31.4	−12.9	−44.3	2	1.0×10^6	−42.7	8.4	−34.3
$d(G_3AG_3CGCTG_3AG_2AG_3)$[c]	K^+	1	1.5×10^7	−12.6	−28.4	−41.0	2	1.0×10^6	−33.1	−1.2	−34.3
$d[TAG_3(T_2AG_3)_3]$[d]	K^+	1	2.0×10^6	−5.0	−30.9	−35.9	3	5.0×10^5	−2.0	−30.5	−32.5
$d(T_2AG_3)_4$[d]	K^+	1	2.0×10^6	−9.4	−26.5	−35.9	3	5.0×10^5	−6.0	−26.5	−32.5
$d[AG_3(T_2AG_3)_3\ T]$[d]	K^+	1	5.0×10^8	−5.2	−44.4	−49.6	3	4.0×10^6	−2.2	−35.4	−37.7
$d[AG_3(T_2AG_3)_3TT]$[d]	K^+	1	5.0×10^8	−7.9	−41.7	−49.6	3	2.0×10^6	−3.3	−32.6	−35.9
$d[TAG_3(T_2AG_3)_3]$[d]	K^+ + PEG	1	3.0×10^8	−30.7	−17.9	−48.6	3	1.0×10^7	−5.0	−35.7	−40.7
$d(T_2AG_3)_4$[d]	K^+ + PEG	1	2.0×10^8	−26.0	−21.4	−47.4	3	6.0×10^6	−6.9	−32.0	−38.9
$d[AG_3(T_2AG_3)_3\ T]$[d]	K^+ + PEG	1	2.0×10^8	−65.2	17.8	−47.4	3	4.0×10^6	−12.3	−25.8	−38.1
$d[AG_3(T_2AG_3)_3TT]$[d]	K^+ + PEG	1	1.0×10^8	−102.5	55.6	−46.9	3	7.0×10^6	−17.4	−21.8	−39.2
$[d(TAG_3)_2]_2$[e]	K^+	2	7.0×10^5	−13.8	−19.7	−33.5	–	–	–	–	–

[a] n_1 and n_2 refer to the number of ligand molecules interacting in each binding event; the final stoichiometry of the complexes is $n_{tot} = n_1 + n_2$

[b] From [52]

[c] From [60]

[d] From [16]

[e] From [73]

with the terminal G-tetrad, whereas the weaker binding was attributable to an external binding mode of TMPyP4, probably with the loops of G-quadruplex structures. Interestingly, they found that the TMPyP4 showed preferential binding to parallel *c-myc* and *c-kit* G-quadruplexes (10^7 M^{-1}) over the antiparallel telomeric one (10^6 M^{-1}) and its interaction with duplex DNA was one order of magnitude lower than G-quadruplex interaction. That study underlined that the differences in the loop orientation, originated from different G-quadruplex conformations, address the binding of TMPyP4 and other small molecules, thus playing a key role in molecular recognition processes.

A few years ago Xue et al. showed that polyethylene glycol (PEG) induces a structural transition of human telomeric G-quadruplexes to a parallel-stranded structure in K^+ solution [63], the same structure adopted in the crystalline state [64]. In the last few years PEG was believed to mimic molecular crowding conditions, close to those of cellular environment, even though, recently, a number of questions have arisen about the actual mechanism of PEG action [65–68]. In a recent paper, Wei et al. studied the interaction of the d[AGGG(TTAGGG)$_3$] sequence with TMPyP4 in PEG 200-containing buffer [69]. They confirmed that the PEG 200 induces the formation of a parallel-stranded G-quadruplex structure and found two independent binding sites for the TMPyP4. The binding mode of TMPyP4 to G-quadruplexes therefore depends on the environment used in the experiments, and also on the structure and composition of the host DNA. Indeed, recent studies described the use of CD and ITC for a characterization of the interaction between TMPyP4 and G-quadruplexes formed by four different human telomeric sequences with 5′-flanking or 3′-flanking bases, in different solution conditions (Table 1) [16]. Two sequences with 5′-flanking thymines assemble to form an antiparallel conformation known as hybrid-1 as the major conformation. The other two sequences with 3′-flanking bases are also structured in an antiparallel conformation, but with different loop arrangements, known as hybrid-2 [70–72]. CD spectroscopy allowed one to follow the TMPyP4 binding since the ligand addition promoted remarkable changes in the CD spectra, due to conformational rearrangement of the G-quadruplex structures, probably from the hybrid structures into a basket one. CD spectra were recorded at different TMPyP4/G-quadruplex molar ratios to calculate the fraction of ligand bound (α_b) (see Sect. 2.2.3). Binding curves were obtained by plotting the α_b as a function of the molar ratio (Fig. 6b). The results of the interpolation analysis suggested that the investigated G-quadruplexes are able to bind up to four TMPyP4 molecules. The K_b values, also obtained from the interpolation analysis, were indicative of the global events and they took into account both binding and conformational changes. The thermodynamic analysis clearly indicated that the sequences with 3′-flanking bases d[AGGG(TTAGGG)$_3$T] and d[AGGG(TTAGGG)$_3$TT], forming the hybrid-2-structures, had a tenfold higher affinity (K_b ~10^6 M^{-1}) for TMPyP4 compared to the sequences d[TAGGG(TTAGGG)$_3$] and d(TTAGGG)$_4$, forming the hybrid-1 structures, (K_b ~10^5 M^{-1}).

On the other hand, in a solution containing 40% PEG 200, the addition of TMPyP4 did not change the shape of the CD spectra of G-quadruplexes, indicating

that, under those conditions, the TMPyP4 interaction did not cause conformational changes in the parallel-stranded G-quadruplex structure. This did not allow the determination of a binding curve. The same occurred for the TMPyP4-duplex interaction. Indeed, for example, TMPyP4 addition did not change the shape of the spectra of duplex DNA having self-complementary CGCGAATTCGCG sequence, which was utilized to check the selectivity. That finding suggested that the TMPyP4 interaction did not promote any conformational change in the duplex DNA structure and that CD methodology cannot be used to obtain thermodynamic parameters under these conditions.

Conversely, the calorimetric analysis of the porphyrin interaction with both G-quadruplexes and duplex allowed one to obtain a complete energetic picture in both PEG-free and PEG-containing solutions (Table 1). In both solution conditions, ITC revealed that the G-quadruplex binding process comprises two sequential events, the first in which one TMPyP4 molecule interacts with the G-quadruplex structures and the second in which three other molecules bind to the structures (Fig. 6c), thus leading to the formation of a final complex composed of one molecule of G-quadruplex and four TMPyP4 molecules, as observed by CD experiments.

In PEG-free solutions, both binding events cooperated to promote a conformational change from the hybrid structures to a basket-type structure, according to the spectroscopic-derived binding curves. The calorimetric results showed that the global Gibbs energy change for the binding of TMPyP4 to the hybrid-1 structures is lower (about -68 kJ mol^{-1}) than that for the hybrid-2 structures (about -86 kJ mol^{-1}). This difference could arise from the higher stability of the hybrid-1 structures due to the presence of an adenine cap that would be destroyed as a result of the interaction-induced conformational change [72]. The thermodynamic signature of the two binding events (Fig. 6c, inset bar graph) for each G-quadruplex sample suggests that the TMPyP4 interaction is entropically driven, and the small enthalpic contribution (ΔH°) acts synergistically with the larger entropic one ($-T\Delta S^{\circ}$), clearly indicating that the driving force of the binding process is not represented by the formation of new interactions.

In PEG-containing solution the thermodynamic characterization of the binding events (Table 1) did not show significant differences between the four sequences and revealed that the two binding events were driven by different thermodynamic contributions. The first binding event was the strongest one and it was enthalpically driven, whereas the second weaker event was entropically driven. Those data clearly suggested that the two consecutive binding events occurred by distinct mechanisms, and they probably involved different regions of the G-quadruplex structure. In particular, it was hypothesized that one of the two terminal G-tetrad planes could be the target region involved in the first event, as for the binding of TMPyP4 to a *c-myc* G-quadruplex determined by NMR. In contrast, the second binding event was entropically driven and, since it occurs with three different molecules of TMPyP4 simultaneously binding to the G-quadruplex, it was suggested that the interacting regions of the G-quadruplex structure could be the three interconnecting loops. This binding mode was found in the crystal structure of another G-quadruplex-TMPyP4 complex [58]. The high entropic contribution

could result from the displacement of water molecules from the surface of the G-quadruplex upon ligand binding.

ITC measurements were also performed on a duplex DNA in both solution conditions [16]. In both cases the results of the calorimetric titrations have shown an exothermic interaction ($\Delta H^\circ = -54$ kJ mol^{-1}) with a stoichiometry of one molecule of TMPyP4 per duplex and similar binding constants ($K_b = 8 \times 10^5$ M^{-1} and $K_b = 6 \times 10^5$ M^{-1} with and without PEG, respectively). Interestingly, in PEG solution the porphyrin was shown to bind the G-quadruplex structure with higher affinity compared to the duplex DNA structure.

Recently, the interaction of Zn(II) 5,10,15,20-tetrakis(*N*-methyl-4-pyridyl)porphyrin (ZnTMPyP4) with a G-quadruplex of sequence d(TAGGG)$_2$ has also been studied by UV–vis and ITC titration experiments [73]. Interestingly, ZnTMPyP4 promotes folding of d(TAGGG)$_2$ into bimolecular parallel G-quadruplex. The binding constants were ~10^6 and ~10^5 M^{-1} from UV and ITC titration experiments, respectively, whereas the binding stoichiometry was found to be 2:1 in both methodologies. On the basis of the thermodynamic signature, inconsistent with an intercalative binding, the authors suggested an external binding mode, specifically an end-stacking interaction. Similar binding interaction between ZnTMPyP4 and an intramolecular G-quadruplex of sequence d[AGGG(TTAGGG)$_3$] was also reported in another work [74]. In that paper an end-stacking mode was also suggested. UV–vis binding experiments were also performed for TMPyP4 with the G-quadruplex [d(TAGGG)$_2$]$_2$. The binding constant was ~10^6 M^{-1}, the same order of magnitude of ZnTMPyP4-[d(TAGGG)$_2$]$_2$ interaction, but the stoichiometry was 3:1, revealing the presence of an additional binding site.

In conclusion, as already described at the beginning of this paragraph, the stoichiometry of TMPyP4 binding to human telomeric G-quadruplexes varies between two and four and the binding constants vary in the range of three orders of magnitude depending not only on DNA sequences and solution conditions but also on the experimental methodologies employed to study the binding process.

A discussion of the binding properties of the multimeric structure formed by the sequence (TTAGGG)$_8$TT with cationic TMPyP4 can be found in [87].

3.2 The Three Side-Chained Triazatruxene Derivative: Azatrux

5,10,15-Tris[4(1-piperidino)butyl]diindolo[3,2-a:3′,2′-c]carbazole (azatrux) is a three side-chained triazatruxene derivative, appropriately designed to bind G-quadruplex structures in an end-stacking mode with its aromatic core (Fig. 7a). Azatrux has proved to be a particularly interesting G-quadruplex ligand, because it is able to inhibit cell growth on cancer cell lines in vitro [75]. A molecular modeling study of the complex effectively showed that azatrux can stack over the terminal G-tetrad plane with the three, positively charged, lateral chains interacting with the DNA phosphates [75]. The knowledge of the energetics that regulates its interaction

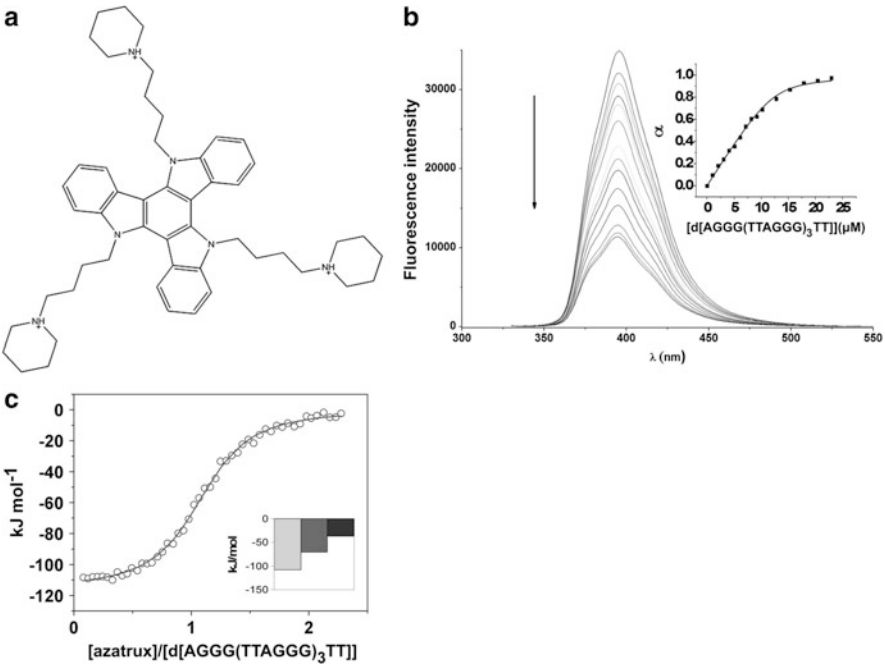


Fig. 7 (**a**) Chemical structure of azatrux. (**b**) Azatrux fluorescence spectra in the absence and presence of successive additions of G-quadruplex at 25 °C. The *arrow* indicates the increasing DNA concentration. The *inset* shows the plot of the fraction of azatrux bound (α) vs DNA concentration, the *black squares* represent the experimental data; the line is the theoretical curve obtained on the basis of an independent and equivalent-sites model. (**c**) ITC binding isotherm for titration of d[AGGG(TTAGGG)$_3$TT] G-quadruplex with azatrux in PEG-containing solution. The *open circles* are the experimental data obtained by integrating the raw data and subtracting the heat of ligand dilution into the buffer. The line is the best fit obtained with the appropriate model. The *inset* shows the thermodynamic signature of the interaction. The enthalpic (ΔH°) and entropic ($T\Delta S^\circ$) contributions are shown in *light gray* and *gray*, respectively; the Gibbs energy change (ΔG°_{298K}) is shown in *dark gray*

with G-quadruplex DNA enhances the opportunity of providing additional information for the optimization of such potential therapeutic agent.

Since it was not possible to characterize the binding of azatrux to the human telomeric G-quadruplex d[AGGG(TTAGGG)$_3$TT] by ITC because of small heats of binding, the binding parameters of the interaction were determined by fluorescence titration experiments [27]. Figure 7b shows the changes in the fluorescence spectrum of azatrux on increasing d[AGGG(TTAGGG)$_3$TT] concentration. Azatrux binds to the G-quadruplex with a 1:1 stoichiometry and with a binding constant of ~5.0×10^5 M^{-1}. CD spectra in the range of ICD signal were also collected. ICD peaks were negative and small and decreased on increasing DNA concentration. This behavior was attributed to an intercalative binding mode or an end-stacking binding mode to G-quadruplex, in agreement with previous studies [44, 45]. In addition, a strong hypochromicity and a drastic fluorescence change

were observed upon azatrux binding to G-quadruplex, suggesting the presence of extended stacking interaction of the azatrux aromatic core, as established for a direct interaction of the drug with a G-tetrad aromatic system.

In a different study, a combination of experimental and computational techniques was employed to characterize the binding of azatrux to G-quadruplex formed by the d[AGGG(TTAGGG)$_3$TT] human telomeric sequence in 40% PEG solution [28]. In that paper, the fruit of collaboration among various groups, the azatrux binding mode to the parallel human telomeric G-quadruplex (conformation induced by high PEG concentrations [63]), was investigated by CD, UV, and fluorescence spectroscopy and the energetics of the binding was characterized by means of ITC. The dynamic and structural features of the azatrux-DNA complex were explored by molecular modeling methods. Further, the binding of azatrux to the parallel human telomeric G-quadruplex was compared with binding to the "standard" parallel G-quadruplex formed by the d(TGGGGT) sequence. Furthermore, the selectivity of azatrux for the human telomeric G-quadruplex relative to another biologically relevant G-quadruplex-forming sequence, such as c-Kit87up, and to duplex DNA was also evaluated in molecular crowding conditions.

The combination of fluorescence, UV, and CD titration experiments showed that one azatrux molecule binds to the d[AGGG(TTAGGG)$_3$TT] without causing conformational change in the G-quadruplex, suggesting an "external" binding mode rather than an intercalative binding mode. Particularly, the large hypochromic shift, observed in the azatrux UV spectrum upon G-quadruplex titration, suggests a binding mode that involves extended stacking interactions between the azatrux aromatic core and G-quadruplex bases. ITC measurements allowed one to gain information on enthalpy and entropy contributions to the energetics of azatrux-d[AGGG(TTAGGG)$_3$TT] interaction. ITC data showed a large enthalpy change (-108 kJ mol^{-1}) and a negative entropic contribution upon azatrux binding (Fig. 7c). The large enthalpy change was attributed to the interaction of positively charged nitrogen atoms of two piperidine side-chains with the negatively charged phosphodiester backbone. A further enthalpy contribution is due to the stacking interaction between the azatrux core and the G-quartet surface. On the other hand, azatrux interaction confers rigidity to the interacting sugar-phosphate backbone in the complex, compared to the free G-quadruplex. This is consistent with the negative observed entropic contribution.

As mentioned above, the binding of azatrux to the "standard" parallel G-quadruplex [d(TGGGGT)]$_4$ was followed by ITC experiments in the presence and absence of PEG. In both cases, the stoichiometry of the azatrux-DNA complex was 2:1, confirming the stoichiometry obtained by NMR titration experiments [28]. On the other hand, the thermodynamic parameters and the binding constant are different in the absence and presence of PEG. In particular, the binding event is enthalpically driven in the presence of PEG and entropically driven in its absence. In the presence of PEG, the binding free energy change is more favorable and the corresponding binding constant is an order of magnitude higher than in the absence of PEG, indicating that the presence of cosolute increases ligand affinity for the G-quadruplex structure. Since PEG does not affect the parallel fold of the

$[d(TGGGGT)]_4$ G-quadruplex, as the CD spectrum does not change, this effect can be directly correlated with the reduced water activity in PEG solution. The comparison of the binding constants and free energies obtained in PEG-containing solutions for the $d[AGGG(TTAGGG)_3TT]$ and $[d(TGGGGT)]_4$ G-quadruplexes, revealed that the azatrux bound these two G-quadruplex structures with similar affinity. The binding was enthalpically driven for both the G-quadruplexes, suggesting similar binding mode, although azatrux bound with a 2:1 stoichiometry the $[d(TGGGGT)]_4$ G-quadruplex and with a 1:1 stoichiometry the human telomeric G-quadruplex. The different stoichiometry was attributed to differences in the geometries of the two G-quadruplex ends, specifically in the human telomeric G-quadruplex the 5′-end and 3′-end are not equivalent and, as suggested by molecular modeling studies, the binding of azatrux could occur only to the 3′-end.

Additionally, the same set of UV–vis and fluorescence titration experiments performed with the human telomeric G-quadruplex was carried out on the $[d(TGGGGT)]_4$ G-quadruplex both in the presence and absence of PEG. The changes in the UV and fluorescence spectra of azatrux upon addition of $[d(TGGGGT)]_4$ were very similar to those observed for the human telomeric sequence, confirming the 2:1 stoichiometry found in the ITC experiments. In UV experiments, the binding constant was $4 \times 10^6\ M^{-1}$ both in the absence and the presence of PEG. In fluorescence experiments, the binding constants were 9×10^6 and $4 \times 10^6\ M^{-1}$ in the absence and in the presence of PEG, respectively.

Selectivity of azatrux for human telomeric G-quadruplex was also evaluated by studying its affinity for the DNA duplex and for the biologically relevant c-Kit87up G-quadruplex by means of UV and fluorescence [76]. Only in the case of the G-quadruplex was there found a measurable affinity with a stoichiometry of 2:1, but a weaker interaction ($K_b = 5 \times 10^{-5}\ M^{-1}$) with respect to that found for the telomeric G-quadruplex.

In conclusion, the whole set of the experimental and computational data suggested that azatrux binds by end-stacking to parallel G-quadruplex structures with higher affinity compared to hybrid structures and that both planar aromatic surface and side chains are important structural elements in optimizing this preferential binding mode. Specifically, computational studies illustrated the key role of the polarity of the side chains in stabilizing the G-quadruplex-azatrux complex by electrostatic interactions.

4 Energetics of Groove Binding G-Quadruplex Ligands

4.1 Distamycin and Its Analogs

Distamycin A (Dst) (Fig. 8) is a well-known minor-groove binder of the duplex DNA, whose stoichiometry of interaction is sequence-dependent [77–79]. However, this molecule and some of its analogs have also been shown to interact with

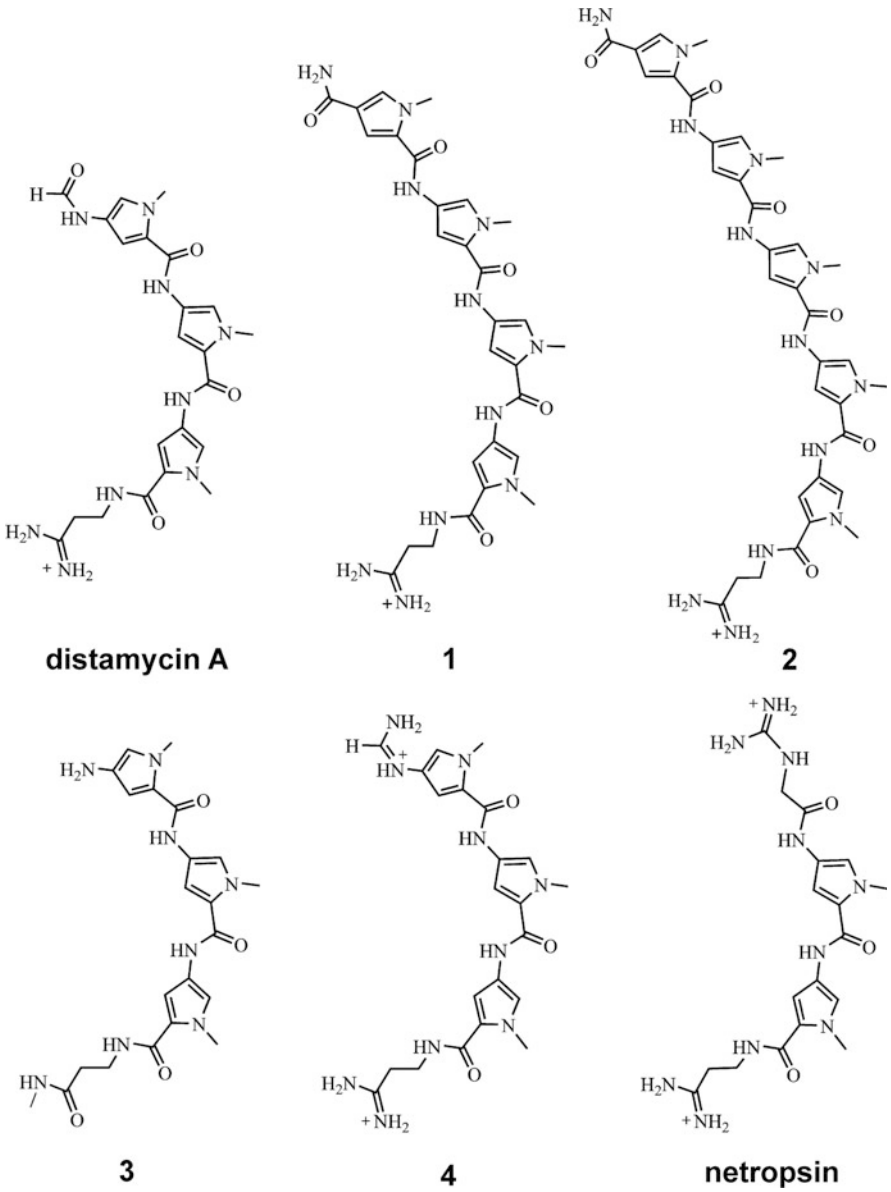


Fig. 8 Chemical structures of distamycin A and its investigated derivatives

DNA quadruplex structures. Additionally, derivatives of Dst have been shown to be effective inhibitors of the human telomerase [80]. Interaction of Dst with tetramolecular G-quadruplex $[d(TGGGGT)]_4$ in K^+ solution shows an interesting ITC binding profile with two distinct binding events, each involving two drug molecules, to give a final 4:1 complex (Fig. 9) [81]. Analysis of the thermodynamic

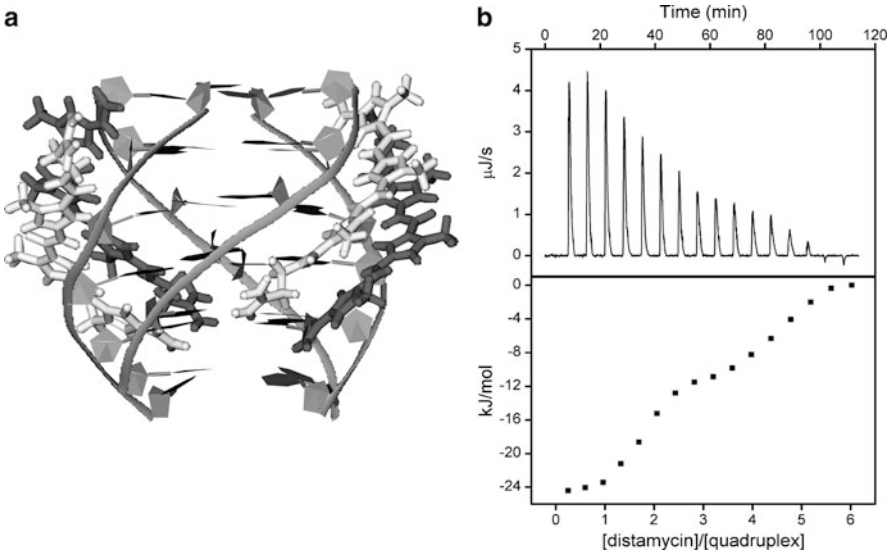


Fig. 9 (**a**) NMR structure of the 4:1 distamycin:[d(TGGGGT)]$_4$ complex. (**b**) ITC data for the binding of distamycin to [d(TGGGGT)]$_4$. The *top panel* shows the raw data; the *bottom panel* shows the data obtained by integrating the peaks and subtracting the heat of ligand dilution into the buffer

Table 2 Thermodynamic parameters for the interaction of distamycin and its investigated derivatives with [d(TGGGGT)]$_4$ obtained by ITC at 25 °C

Ligand	Ion		n^a	K_b (M^{-1})	$\Delta H°$ (kJ mol^{-1})	$-T\Delta S°$ (kJ mol^{-1})	$\Delta G°$ (kJ mol^{-1})	Ref.
Distamycin	K^+	1° binding event	2	4.0×10^5	−8.0	−24	−32	[81]
		2° binding event	2	4.0×10^6	−10.0	−27	−37	
Distamycin	Na^+		1	2.0×10^5	−14.0	−16	−30	[61]
1	K^+		2	2.0×10^6	7.0	−43	−36	[61]
1	Na^+		1	2.3×10^6	10.0	−46	−36	[61]
3	K^+		4	9.0×10^5	6.0	−40	−34	[82]
4	K^+		2	1.9×10^6	−11.0	−25	−36	[83]
Netropsin	K^+		2	1.2×10^6	−10.6	−23	−34	[83]

[a]*n* refers to the number of ligand molecules interacting in each binding event

parameters showed an entropically driven process and suggested an interaction with the grooves of the [d(TGGGGT)]$_4$ quadruplex. NMR study confirmed that four Dst molecules bind, as antiparallel dimers, two opposite G-quadruplex grooves by adopting a crescent shape and by establishing H-bonds with the guanine bases (Fig. 9). NMR also reveals strong coulombic interactions between positively charged amidinium moiety of Dst and the backbone phosphate groups of the G-quadruplex. From the energetic point of view the binding was entropically driven

and the affinity between Dst and $[d(TGGGGT)]_4$ was enhanced (~tenfold) when the ratio drug/G-quadruplex was increased (Table 2).

The importance of the crescent shape extension was investigated by varying the pyrrole units number in Dst [61]. The attention was focused on the interaction of two carbamoyl analogues of Dst (**1** and **2**, Fig. 8), containing four and five pyrrole units, respectively. Experiments revealed that the presence of one additional pyrrole unit affects the affinity as well as the stoichiometry (2:1 ligand/quadruplex) of the binding, whereas, the addition of two pyrrole units leads to a total loss of interaction between the derivative and the $[d(TGGGGT)]_4$. The interactions of Dst, **1** and **2** to the $[d(TGGGGT)]_4$ target were also investigated using a different buffered solution containing Na^+ at fixed ionic strength. ITC experiments revealed that Dst and compound **1** bound to the investigated G-quadruplex in that solution condition, but that compound **2** had a poor binding affinity. However, the presence of Na^+ cation in solution affected the stoichiometry and thermodynamics of the interactions. Specifically, the energetics of binding for the compound **1** was entirely entropically driven (Table 2).

The importance of the unique positive charge of Dst in the interaction with $[d(TGGGGT)]_4$ was investigated by studying an uncharged analog, in which the major change into the ligand structure was the replacement of the amidinium group by an *N*-methyl amide moiety (**3**, Fig. 8) [82]. As for Dst, the binding stoichiometry was found to be 4:1 (ligand/quadruplex), but the thermodynamic parameters determined by ITC for the interaction of **3** were slightly different from those of Dst to the same target in K^+ solution. In both cases, the binding reaction was an entropically driven process. However, the binding of **3** showed a small unfavorable enthalpy change, while a small favorable enthalpy change was observed in the case of Dst. This difference was rationalized by the structural features of the two ligands. Indeed, the positively charged amidinium moiety of Dst interacted with the phosphate groups of the quadruplex, providing a favorable, although small, enthalpy contribution. On the other hand, in the derivative **3** the amidinium group was replaced by an uncharged moiety that could not give this contribution.

The energetics of binding of two dicationic derivatives of Dst (**4** and netropsin, Fig. 8) with $[d(TGGGGT)]_4$ was also characterized by ITC [83]. The experiments revealed that both **4** and netropsin bound to the investigated quadruplex. The thermodynamic profiles of the two ligand-quadruplex interactions were qualitatively similar and the stoichiometry observed was 2:1 (ligand/quadruplex) in both cases (Table 2). The thermodynamic parameters determined by ITC indicated that, similarly to Dst, the association reactions of both ligands with $[d(TGGGGT)]_4$ were entropically driven processes with a small favorable enthalpic contribution, suggesting a similar groove binding mode also in these cases. Interestingly, the 2:1 (ligand/quadruplex) stoichiometry suggested that **4** and netropsin were not able to bind the quadruplex in a dimeric form, most probably due to the doubly charged nature of the molecules that prevented a side-by-side arrangement into the grooves.

The ITC data alone do not establish with certainty the binding mode of these ligands, but, by combining calorimetric and structural data, it is reasonable to assess that the binding into the grooves of G-quadruplexes is characterized by a large

favorable increase of entropy and a small (favorable or unfavorable) enthalpy change, similar to what happens for the binding of a groove binder to duplex DNA [48].

5 Conclusions and Perspectives

The search for best-in-class G-quadruplex ligands, acting on both telomeric and nontelomeric G-quadruplexes, occupies many researchers around the world. Among the plethora of putative G-quadruplex ligands, only a fluoroquinolone derivative, quarfloxin, had entered into the phase II clinical trials for the treatment of carcinoid/neuroendocrine tumors [84]. Unfortunately, quarfloxin was withdrawn from further trials owing to problems with bioavailability. Although the increasing discovery of G-quadruplexes in many gene promoter regions opens new routes of therapeutic targets in oncology, it is still a long way off to being able to obtain G-quadruplex ligands as successful anticancer agents [85]. Physico-chemical properties and structural characteristics of G-quadruplexes of many gene promoter regions make them good targets for a range of drugs. Further, their structural diversity is promising for a high degree of selectivity. Indeed, drug selectivity represents a potential problem for the development of ligands able to bind and stabilize only a specific G-quadruplex structure. Furthermore, the identification of the specific proteins that bind to G-quadruplex forming-sequences modulating gene expression raises a question on the opportunity of studying how the drugs exert their biological effects in the presence of G-quadruplex-protein complexes. Finally, one last consideration on the process of first-in-class drug identification should be carried out. Usually this process is mainly and unquestionably focused on the common drug optimizing paradigm of simultaneously improving selectivity, solubility, and pharmacokinetic properties rather than analyzing binding energetics. Careful investigations on G-quadruplex-ligand energetics can surely help the process of development of G-quadruplex-binding drug, giving quantitative information on the investigated interactions. Fortunately, physico-chemical methodologies suitable to study these aspects, some of which are described in this chapter, are available in many research laboratories.

Acknowledgments The authors would like to thank Dr. Jussara Amato for critical reading of the manuscript. This work was supported by the Italian "Ministero dell'Istruzione, dell'Università e della Ricerca" PRIN 2009 grant; and by the "Associazione Italiana per la Ricerca sul Cancro" [MFAG n° 11947 to B.P.]. The COST Action MP0802 is gratefully acknowledged for offering a stimulating environment for discussions on the topic.

References

1. Blackburn EH (1991) Structure and function of telomeres. Nature 350:569–573
2. de Lange T (2005) Shelterin: the protein complex that shapes and safeguards human telomeres. Genes Dev 19:2100–2110
3. Shay J, Wright W (2006) Telomerase therapeutics for cancer: challenges and new directions. Nat Rev Drug Discov 5:577–584
4. Zahler AM, Williamson JR, Cech TR et al (1991) Inhibition of telomerase by G-quartet DNA structures. Nature 350:718–720
5. Blackburn EH (1992) Telomerases. Annu Rev Biochem 61:113–129
6. Kim NW, Piatyszek MA, Prowse KR et al (1994) Specific association of human telomerase activity with immortal cells and cancer. Science 266:2011–2015
7. De Cian A, Lacroix L, Douarre C et al (2008) Targeting telomeres and telomerase. Biochimie 90:131–155
8. Neidle S (2010) Human telomeric G-quadruplex: the current status of telomeric G-quadruplexes as therapeutic targets in human cancer. FEBS J 277:1118–1125
9. Freire E (2008) Do enthalpy and entropy distinguish first in class from best in class? Drug Discov Today 13:869–874
10. Ladbury JE, Klebe G, Freire E (2010) Adding calorimetric data to decision making in lead discovery: a hot tip. Nat Rev Drug Discov 9:23–27
11. Meanwell NA (2011) Improving drug candidates by design: a focus on physicochemical properties as a means of improving compound disposition and safety. Chem Res Toxicol 24:1420–1456
12. Doyle ML (1997) Characterization of binding interactions by isothermal titration calorimetry. Curr Opin Biotechnol 8:31–35
13. Ababou A, Ladbury JE (2007) Survey of the year 2005: literature on applications of isothermal titration calorimetry. J Mol Recognit 20:4–14
14. Pagano B, Mattia CA, Giancola C (2009) Applications of isothermal titration calorimetry in biophysical studies of G-quadruplexes. Int J Mol Sci 10:2935–2957
15. Pagano B, Giancola C (2007) Energetics of quadruplex-drug recognition in anticancer therapy. Curr Cancer Drug Targets 7:520–540
16. Martino L, Pagano B, Fotticchia I et al (2009) Shedding light on the interaction between TMPyP4 and human telomeric quadruplexes. J Phys Chem B 113:14779–14786
17. Peng D, Tan JH, Chen SB et al (2010) Bisaryldiketene derivatives: a new class of selective ligands for c-myc G-quadruplex DNA. Bioorg Med Chem 18:8235–8242
18. Bhadra K, Kumar GS (2011) Interaction of berberine, palmatine, coralyne, and sanguinarine to quadruplex DNA: a comparative spectroscopic and calorimetric study. Biochim Biophys Acta 1810:485–496
19. Chaires JB (2008) Calorimetry and thermodynamics in drug design. Annu Rev Biophys 37:135–151
20. Leavitt S, Freire E (2001) Direct measurement of protein binding energetics by isothermal titration calorimetry. Curr Opin Struct Biol 11:560–566
21. Brown A (2009) Analysis of cooperativity by isothermal titration calorimetry. Int J Mol Sci 10:3457–3477
22. LaKowicz JR (2006) Principles of fluorescence spectroscopy, vol 1, 3rd edn. Springer, New York, 954 p
23. Miannay FA, Banyasz A, Gustavsson T et al (2009) Excited states and energy transfer in G-quadruplexes. J Phys Chem C 113:11760–11765
24. Mendez MA, Szalai VA (2009) Fluorescence of unmodified oligonucleotides: a tool to probe G-quadruplex DNA structure. Biopolymers 91:841–850
25. Dao NT, Haselsberger R, Michel-Beyerle ME et al (2011) Following G-quadruplex formation by its intrinsic fluorescence. FEBS Lett 585:3969–3977

26. Chang CC, Chien CW, Lin YH et al (2007) Investigation of spectral conversion of d $(TTAGGG)_4$ and d$(TTAGGG)_{13}$ upon potassium titration by a G-quadruplex recognizer BMVC molecule. Nucleic Acids Res 35:2846–2860
27. Cummaro A, Fotticchia I, Franceschin M et al (2011) Binding properties of human telomeric quadruplex multimers: a new route for drug design. Biochimie 93:1392–1400
28. Petraccone L, Fotticchia I, Cummaro A et al (2011) The triazatruxene derivative azatrux binds to the parallel form of the human telomeric G-quadruplex under molecular crowding conditions: biophysical and molecular modeling studies. Biochimie 93:1318–1327
29. Monchaud D, Allain C, Teulade-Fichou MP (2006) Development of a fluorescent intercalator displacement assay (G4-FID) for establishing quadruplex-DNA affinity and selectivity of putative ligands. Bioorg Med Chem Lett 16:4842–4845
30. Monchaud D, Allain C, Bertrand H et al (2008) Ligands playing musical chairs with G-quadruplex DNA: a rapid and simple displacement assay for identifying selective G-quadruplex binders. Biochimie 90:1207–1223
31. Förster T (1948) Zwischenmolekulare Energiewanderung und Fluoreszenz. Ann Phys 437:55–75
32. Mergny JL, Maurizot JC (2001) Fluorescence resonance energy transfer as a probe for G-quartet formation by a telomeric repeat. Chembiochem 2:124–132
33. De Cian A, Guittat L, Kaiser M et al (2007) Fluorescence-based melting assays for studying quadruplex ligands. Methods 42:183–195
34. Juskowiak B (2006) Analytical potential of the quadruplex DNA-based FRET probes. Anal Chim Acta 568:171–180
35. Fasman GD (1996) Circular dichroism and the conformational analysis of biomolecules. Plenum Press, New York
36. Masiero S, Trotta R, Pieraccini S et al (2010) A non-empirical chromophoric interpretation of CD spectra of DNA G-quadruplex structures. Org Biomol Chem 8:2683–2692
37. Karsisiotis AI, Hessari NM, Novellino E et al (2011) Topological characterization of nucleic acid G-quadruplexes by UV absorption and circular dichroism. Angew Chem Int Ed Engl 50:10645–10648
38. Li W, Wu P, Ohmichi T et al (2002) Characterization and thermodynamic properties of quadruplex/duplex competition. FEBS Lett 526:77–81
39. Li J, Correia JJ, Wang L et al (2005) Not so crystal clear: the structure of the human telomere G-quadruplex in solution differs from that present in a crystal. Nucleic Acids Res 33:4649–4659
40. Balagurumoorthy P, Brahmachari SK, Mohanty D et al (1992) Hairpin and parallel quartet structures for telomeric sequences. Nucleic Acids Res 20:4061–4067
41. Jin R, Gaffney BL, Wang C et al (1992) Thermodynamics and structure of a DNA tetraplex: a spectroscopic and calorimetric study of the tetramolecular complexes of d(TG_3T) and d$(TG_3T_2G_3T)$. Proc Natl Acad Sci USA 89:8832–8836
42. Lu M, Guo Q, Kallenbach NR (1993) Thermodynamics of G-tetraplex formation by telomeric DNAs. Biochemistry 32:598–601
43. Sun H, Tang Y, Xiang J et al (2006) Spectroscopic studies of the interaction between quercetin and G-quadruplex DNA. Bioorg Med Chem Lett 16:3586–3589
44. Garbett NC, Ragazzon PA, Chaires JB (2007) Circular dichroism to determine binding mode and affinity of ligand-DNA interactions. Nat Protoc 2:3166–3172
45. Hudson JS, Brooks SC, Graves DE (2009) Interactions of actinomycin D with human telomeric G-quadruplex DNA. Biochemistry 48:4440–4447
46. Arora A, Balasubramanian C, Kumar N et al (2008) Binding of berberine to human telomeric quadruplex – spectroscopic, calorimetric and molecular modeling studies. FEBS J 275:3971–3983
47. Chen Z, Zheng KW, Hao YH et al (2009) Reduced or diminished stabilization of the telomere G-quadruplex and inhibition of telomerase by small chemical ligands under molecular crowding condition. J Am Chem Soc 131:10430–10438

48. Chaires JB (2006) A thermodynamic signature for drug-DNA binding mode. Arch Biochem Biophys 453:26–31
49. Zhang HJ, Wang XF, Wang P et al (2008) Spectroscopic study on the binding of a cationic porphyrin to DNA G-quadruplex under different K+ concentrations. Photochem Photobiol Sci 7:948–955
50. Wheelhouse RT, Sun D, Han H et al (1998) Cationic porphyrins as telomerase inhibitors: the interaction of tetra-(N-methyl-4-pyridyl)porphine with quadruplex DNA. J Am Chem Soc 120:3261–3262
51. Siddiqui-Jain A, Grand CL, Bearss DJ et al (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. Proc Natl Acad Sci USA 99:11593–11598
52. Haq I, Trent JO, Chowdhry BZ et al (1999) Intercalative G-tetraplex stabilization of telomeric DNA by a cationic porphyrin. J Am Chem Soc 121:1768–1779
53. Han FXG, Wheelhouse RT, Hurley LH (1999) Cationic porphyrins as telomerase inhibitors: the interaction of tetra (N-methyl-4-pyridyl) porphyrin with quadruplex DNA. J Am Chem Soc 121:3561–3570
54. Han H, Langley DR, Rangan A et al (2001) Selective interactions of cationic porphyrins with G-quadruplex structures. J Am Chem Soc 123:8902–8913
55. Erra E, Petraccone L, Esposito V et al (2005) Interaction of porphyrin with G-quadruplex structures. Nucleosides Nucleotides Nucleic Acids 24:753–756
56. Phan AT, Kuryavyi V, Gaw HY et al (2005) Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. Nat Chem Biol 1:167–173
57. Wei C, Jia G, Yuan J et al (2006) A spectroscopic study on the interactions of porphyrin with G-quadruplex DNAs. Biochemistry 45:6681–6691
58. Parkinson GN, Ghosh R, Neidle S (2007) Structural basis for binding of porphyrin to human telomeres. Biochemistry 46:2390–2397
59. Gaynutdinov TI, Neumann RD, Panyutin IG (2008) Structural polymorphism of intramolecular quadruplex of human telomeric DNA: effect of cations, quadruplex-binding drugs and flanking sequences. Nucleic Acids Res 36:4079–4087
60. Arora A, Maiti S (2008) Effect of loop orientation on quadruplex-TMPyP4 interaction. J Phys Chem B 112:8151–8159
61. Pagano B, Virno A, Mattia CA et al (2008) Targeting DNA quadruplexes with distamycin A and its derivatives: an ITC and NMR study. Biochimie 90:1224–1232
62. Gray RD, Li J, Chaires JB (2009) Energetics and kinetics of a conformational switch in G-quadruplex DNA. J Phys Chem B 113:2676–2683
63. Xue Y, Kan ZY, Wang Q et al (2007) Human telomeric DNA forms parallel-stranded intramolecular G-quadruplex in K+ solution under molecular crowding condition. J Am Chem Soc 129:11185–11191
64. Parkinson GN, Lee MP, Neidle S (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. Nature 417:876–880
65. Zhou HX, Rivas G, Minton AP (2008) Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. Annu Rev Biophys 37:375–397
66. Miller MC, Buscaglia R, Chaires JB et al (2010) Hydration is a major determinant of the G-quadruplex stability and conformation of the human telomere 3′ sequence of d $[AG_3(TTAG_3)_3]$. J Am Chem Soc 132:17105–17107
67. Hansel R, Lohr F, Foldynova-Trantirkova S et al (2011) The parallel G-quadruplex structure of vertebrate telomeric repeat sequences is not the preferred folding topology under physiological conditions. Nucleic Acids Res 39:5768–5775
68. Petraccone L, Pagano B, Giancola C (2012) Studying the effect of crowding and dehydration on DNA G-quadruplexes. Methods. doi:10.1016/j.ymeth.2012.02.011
69. Wei C, Jia G, Zhou J et al (2009) Evidence for the binding mode of porphyrins to G-quadruplex DNA. Phys Chem Chem Phys 11:4025–4032

70. Phan AT, Luu KN, Patel DJ (2006) Different loop arrangements of intramolecular human telomeric (3+1) G-quadruplexes in K+ solution. Nucleic Acids Res 34:5715–5719
71. Dai J, Carver M, Punchihewa C et al (2007) Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K+ solution: insights into structure polymorphism of the human telomeric sequence. Nucleic Acids Res 35:4927–4940
72. Dai J, Punchihewa C, Ambrus A et al (2007) Structure of the intramolecular human telomeric G-quadruplex in potassium solution: a novel adenine triple formation. Nucleic Acids Res 35:2440–2450
73. Bhattacharjee AJ, Ahluwalia K, Taylor S et al (2011) Induction of G-quadruplex DNA structure by Zn(II) 5,10,15,20-tetrakis (N-methyl-4-pyridyl)porphyrin. Biochimie 93:1297–1309
74. Pan J, Zhang S (2009) Interaction between cationic zinc porphyrin and lead ion induced telomeric guanine quadruplexes: evidence for end-stacking. J Biol Inorg Chem 14:401–407
75. Ginnari-Satriani L, Casagrande V, Bianco A et al (2009) A hydrophilic three side-chained triazatruxene as a new strong and selective G-quadruplex ligand. Org Biomol Chem 7:2513–2516
76. Todd AK, Haider SM, Parkinson GN et al (2007) Sequence occurrence and structural uniqueness of a G-quadruplex in the human c-kit promoter. Nucleic Acids Res 35:5799–5808
77. Thomas M, Varshney U, Bhattacharya S (2002) Distamycin analogs without leading amide at the N-terminus: comparative binding properties to AT and GC rich DNA sequences. Eur J Org Chem 3604–3615
78. Bhattacharya S, Thomas M (2000) DNA binding properties of novel distamycin analogs that lack the leading amide unit at the N-terminus. Biochem Biophys Res Commun 267:139–144
79. Bhattacharya S, Thomas M (2000) Facile synthesis of oligopeptide distamycin analogs devoid of hydrogen bond donors or acceptors at the N-terminus: sequence-specific duplex DNA binding as a function of peptide chain length. Tetrahedron Lett 41:5571–5575
80. Zaffaroni N, Lualdi S, Villa R et al (2002) Inhibition of telomerase activity by a distamycin derivative: effects on cell proliferation and induction of apoptosis in human cancer cells. Eur J Cancer 38:1792–1801
81. Martino L, Virno A, Pagano B et al (2007) Structural and thermodynamic studies of the interaction of distamycin A with the parallel quadruplex structure $[d(TGGGGT)]_4$. J Am Chem Soc 129:16048–16056
82. Cosconati S, Marinelli L, Trotta R et al (2010) Structural and conformational requisites in DNA quadruplex groove binding: another piece to the puzzle. J Am Chem Soc 132:6425–6433
83. Pagano B, Fotticchia I, De Tito S et al (2010) Selective binding of distamycin A derivative to G-quadruplex structure [d(TGGGGT)](4). J Nucleic Acids 2010: pii: 247137
84. Drygin D, Siddiqui-Jain A, O'Brien S et al (2009) Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis. Cancer Res 69:7653–7661
85. Balasubramanian S, Hurley LH, Neidle S (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? Nat Rev Drug Discov 10:261–275
86. Randazzo A, Spada GP, Webba da Silva M (2012) Circular dichroism of Quadruplex structures. Top Curr Chem DOI: 10.1007/128_2012_331
87. Petraccone L (2012) Higher-Order quadruplex structures. Top Curr Chem DOI: 10.1007/128_2012_350

Top Curr Chem (2013) 330: 243–274
DOI: 10.1007/128_2012_334

Published online: 3 July 2012

Tetramolecular Quadruplex Stability and Assembly

Phong Lan Thao Tran, Anne De Cian, Julien Gros, Rui Moriyama, and Jean-Louis Mergny

Abstract Guanine quadruplexes (G4) are unusual four-stranded nucleic acid structures formed by G-rich DNA/RNA. Beyond their likely biological relevance, the self-assembly, stability, and rigidity of these structures are also interesting for nanotechnology and biotechnology applications. Therefore, efforts are carried out to understand the rules that govern stability and folding of G-quadruplexes. We focus this chapter on tetramolecular conformations which are simple tractable models. We report here the experimental parameters, molecules, and modifications that affect thermal stability and/or association kinetics of these structures. Some chemical modifications which facilitate tetramolecular quadruplex formation and can be useful for nano- or biotechnology are also described.

Keywords Association kinetics, Chemical modifications, Nanotechnology and biotechnology applications, Tetramolecular quadruplexes, Thermal stability

Contents

1 Introduction 244
2 Quartet Formation 246
2.1 Classical Quartet 246
2.2 Non-G-Quartets 247
2.3 Pentads, Hexads, Heptads, and Octads 248
3 Stability and Association Kinetics of Tetramolecular Quadruplexes 249
3.1 Presentation 249
3.2 Association Process Analysis 249

P.L.T. Tran (✉), R. Moriyama and J.-L. Mergny (✉)
Université de Bordeaux, Laboratoire ARNA, 33000 Bordeaux, France

INSERM, U869, Laboratoire ARNA, IECB, 33600 Pessac, France
e-mail: thao.tran@inserm.fr; jean-louis.mergny@inserm.fr

A. De Cian and J. Gros
INSERM, U565, CNRS UMR7196, Muséum National d'Histoire Naturelle,
75231 Paris Cedex 05, France

3.3 Thermal Stability Measurements 251
4 Stability and Association Kinetic Effectors for Tetramolecular Quadruplexes 251
4.1 Effects of Temperature 251
4.2 Effects of Ionic Strength 253
4.3 Effects of pH 254
4.4 Effects of Sequence and Concentration 254
4.5 Effects of the 5′-Phosphate Group 254
5 Effects of Chaperone Molecules 255
5.1 Effect of Proteins 255
5.2 Effect of Quadruplex Ligands 255
5.3 Effect of Polymers 257
6 Effects of Chemical Modifications 259
6.1 RNA and 2′-Sugar Modified Analogs 259
6.2 Further Backbone Modifications 261
6.3 Guanine Base Modifications 263
7 Potential Folding Pathway of Tetramolecular Quadruplexes 264
8 Biotechnology and Nanotechnology Applications 265
8.1 DNA Nanostructures: G-Wires, Frayed Wires, and "Synapsable" DNA 265
8.2 Quadruplex as Ion Channels 266
8.3 Nanoparticles Connected via Tetramolecular Quadruplexes 266
9 Conclusion 266
References 267

Abbreviations

A	Adenine
C	Cytosine
CD	Circular dichroism
DNA	Deoxyribonucleic acid
G	Guanine
HIV-1	Human immunodeficiency virus type 1
LNA	Locked nucleic acid
NMR	Nuclear magnetic resonance
O-RNA	2′-*O*-Methyl ribonucleic acid
PEG	Polyethylene glycol
PNA	Peptide nucleic acid
RNA	Ribonucleic acid
T	Thymine
U	Uracil
UTR	Untranslated region

1 Introduction

Self-assembly of guanosine 5′-monophosphates (5′-GMP) or guanine-rich poly- and oligonucleotides has been observed for several decades. In 1962, Gellert et al. [1] determined that guanine bases form non-canonical structures based on G-quartets. In a similar manner, G-rich nucleic acids assemble into higher order structures called guanine quadruplexes. G-quadruplex structures exhibit right-handed helicity and

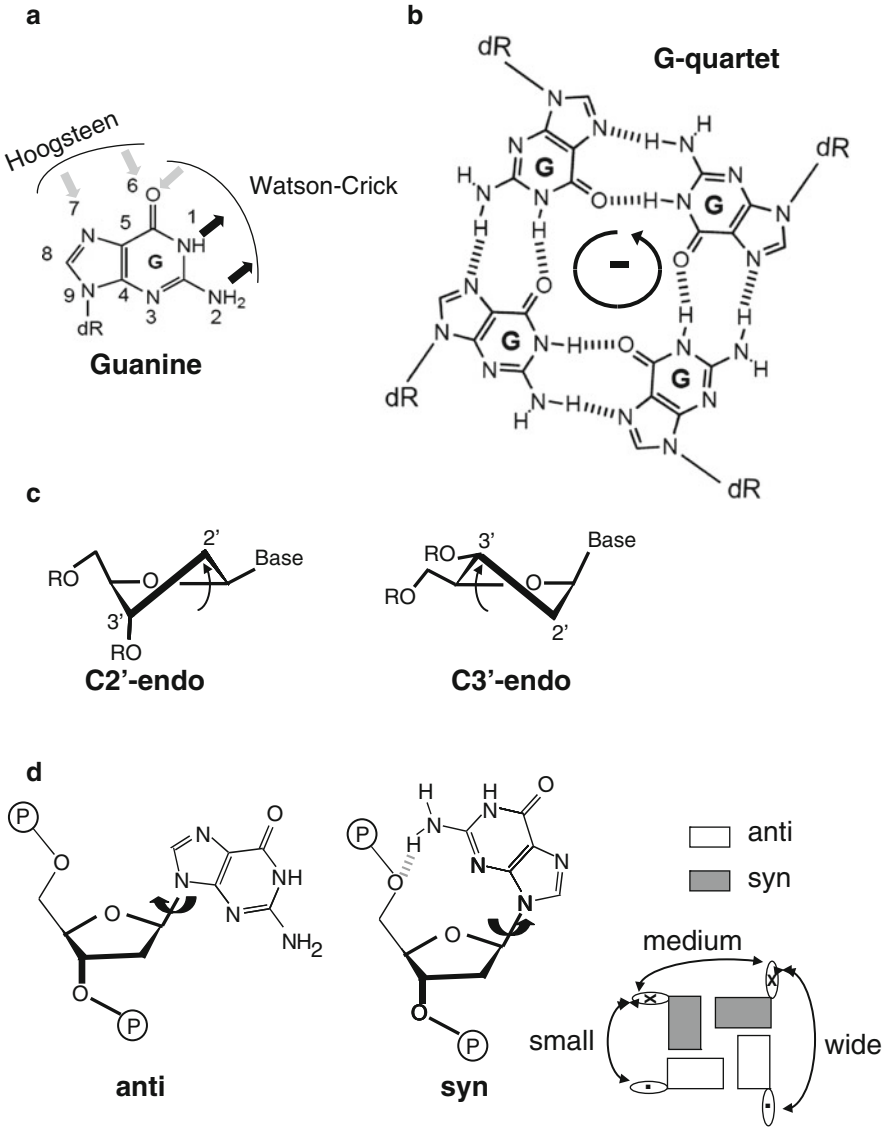


Fig. 1 *Guanine and G-quartet*. (**a**) Chemical formula of canonical guanine and two hydrogen bonding faces (Watson–Crick and Hoogsteen) which are implicated in G-quartet formation. *Arrows* indicate H-bond donors (in *black*) and acceptors (in *gray*). (**b**) Classical G-quartet structure with anticlockwise rotation (−) of the donor NH to the acceptor C=O hydrogen bonds. (**c**) Two most favorable sugar conformations of guanine of quartet: C2′- or C3′-*endo*. (**d**) Two torsion angles of guanine glycosidic bond (*syn* and *anti*) that determine groove dimension of G-quartet: wide, medium and small

result from the hydrophobic stacking of two or more G-quartets, in which four guanines establish a cyclic array of hydrogen bonds from the Watson–Crick and the Hoogsteen faces (Fig. 1a). The stacking of G-quartets is stabilized by cations such as Na^+ or K^+. Bioinformatics studies identified about 376,000 putative

quadruplex sequences (PQS) in the human genome [2]. Such structures may form in specific genome regions such as gene promoters, ribosomal DNA, minisatellites, telomere regions, as well as 5′-UTR, 3′-UTR, or intron regions. Several recent publications in high profile journals indicated an implication of these structures in key biological processes [3–13], making them an attractive target for drug design [4, 14–17]. Besides biological applications of G-quadruplexes, their self-assembly, stability and rigidity properties are of particular interest for the nanotechnology and the biotechnology fields.

It is therefore important to understand the rules that govern the stability and folding kinetics of these complexes. In contrast to bi- and unimolecular conformations, tetramolecular quadruplexes formed by short nucleic acid sequences containing a single guanine tract display two major advantages for this analysis: (1) a limited polymorphism (even if unusual planar base associations can be formed) [18–20] and (2) kinetic inertia that facilitates experimental studies. For this reason, the folding of uni- or bimolecular quadruplexes will not be dealt with in this chapter – just the assembly of tetramolecular quadruplexes.

2 Quartet Formation

2.1 Classical Quartet

A G-quartet is a planar association of four guanines held together by four central H-bonds between H^1 of nitrogen and O^6 of carbonyl group and also by four external H-bonds between H^2 of amine and N^7 (Fig. 1b). A G-quartet has two diastereotopic faces, a “head” (+) and a “tail” (−). The “head” of the G-quartet is defined as the face showing a clockwise orientation of the internal hydrogen bonds, from the donor (NH) to the acceptor (C=O) group, and vice versa for the “tail.”

The nucleoside sugars (ribose or deoxyribose) can adopt several conformations. The two most favorable sugar conformations are C2′-*endo* (found in most DNA G-quadruplexes) and C3′-*endo* (found in most RNA structures) (Fig. 1c) [21].

The guanine bases are covalently linked to the sugar via the glycosidic bond that can display two major different torsion angles: *syn* and *anti* (Fig. 1d). The *anti* conformation can be formed whatever the sugar conformation (C2′- or C3′-*endo*), whereas the *syn* position is unfavorable in the case of C3′-*endo* sugar conformation because of the steric hindrance between O3′ and C5′. Despite this steric effect, the *syn* position is possible for guanine thanks to the H-bond between O5′ and H^2 of amine. Each G-quadruplex presents four grooves defined by two neighboring guanines in a G-quartet. According to their respective *syn* or *anti* conformations, grooves can be wide, medium, or small [22], contributing to the conformational diversity of G-quadruplex structures. However, this diversity is more limited in the case of tetramolecular quadruplexes; when all strands are parallel, all guanine in a

quartet must adopt the same glycosidic bond angle, leading to all-*syn* or all-*anti* quartets (the latter being the general case), and all grooves are of medium size.

The stacking of G-quartets is stabilized by monovalent (Rb^+, NH_4^+, K^+, Na^+, Cs^+, Li^+), divalent (Sr^{2+}, Ca^{2+}, Ba^{2+}, Mg^{2+}), or trivalent (Tb^{3+}, Eu^{3+}) [23, 24] cations according to their ionic radii and their hydration energies. There are indeed two different ways for cations to interact with G-quadruplex structures: (1) cation–dipole interactions through 2×4 central oxygen atoms (inner ions) from two successive G-quartets, reducing the electronic repulsion and (2) electrostatic interactions with the grooves and the phosphate backbone (outer ions). The latter mode is shared with most nucleic acid structures while the first is unique for G-quadruplexes.

The order of cation ability to stabilize and/or to induce G-quadruplex structures is as follow: $K^+ > NH_4^+ > Rb^+ > Na^+ > Cs^+ > Li^+$ for monovalent cations [25] and $Sr^{2+} > Ba^{2+} > Ca^{2+} > Mg^{2+}$ for divalent cations [26]. The residence time of NH_4^+ cations bound between central quartets of a G-quadruplex is generally much shorter than the lifetime of a G-quadruplex. This suggests that cation movement through the G-quadruplex may occur without opening the structure. Most measurements have been performed on bi- or intra-molecular complexes with NH_4^+ cations residence lifetimes in the 139 ms to 1.7 s range [20, 27–30]. G-quadruplex could thus be compared to ion channels, an intrinsic property that could be used for nanotechnology applications (see Sect. 7), i.e., the synthesis of conducting molecular wires or synthetic transmembrane ion transporters.

2.2 Non-G-Quartets

Besides guanines, other canonical nucleobases can also form quartets; either *homo-quartets* formed by the same base (A-, T-, C-, or U-quartets) or *mixed quartets* formed by two different bases (G:C:G:C, G:T:G:T, or A:T:A:T). The design of a tetramolecular complex incorporating a *mixed quartet* is not straightforward, as it necessitates the assembly of two different strands. In contrast, it is simple to design a tetramolecular complex that would lead to the formation of A-, T-, C-, or U- *homo-quartets*; we studied in detail the stability of these G4 structures, bearing in mind that structural methods would be required to determine whether these quartets are actually formed. In general, even if such quartets have been observed by NMR, they do contribute greatly to the stability of the quadruplex, and often have a very detrimental impact on association rate and/or thermal stability. The sequences and the corresponding unusual quartets that can be formed under different ionic conditions are displayed in Table 1. Non-G-quartets, observed by NMR or crystallography, are formed by G-rich sequences containing several G-quartets. These G-quartets stabilize the formation of non-G-quartets which may slow cation movement once presented at the 5′ end [20].

Table 1 Unusual quartets and corresponding sequences described in the literature

	Unusual quartets	Sequences ($5' \rightarrow 3'$)	Condition of formation	References
Homo-quartets	A-quartet	$d(AG_3T)$	K^+	[31]
		$d(^{Br}U^a)$-$r(GAG_2U)$	Na^+	[32]
		$r(U)$-$d(^{Br}G^a)$-$r(AG_2U)$	Na^+	[33]
	T-quartet	$d(TG_4T)$	Na^+, K^+, or NH_4^+	[20]
		$d(TG_3TG_3C)$	K^+	[34]
	C-quartet	$d(TG_3CG_3T)$	Na^+	[35]
	U-quartet	$r(UG_4U)$	K^+ or Sr^{2+}	[36, 37]
		$d(UG_4T)$	Na^+	[20]
		$r(UG_2UGU)$	K^+	[38]
Mixed quartets	G-C quartet	$d(GCG_2X^aG_2Y^a)$	Na^+	[39]
		$d(GCG_2T_3GCG_2)$	Na^+	[40]
		$d(G_3CT_4G_3C)$	Na^+, K^+, or NH_4^+	[41–43]
		$d(G_3ATC)$	K^+	[44]
	G-C or A-T quartets	$d(GCG_2T_2G_2AT)$	Na^+	[45]
		$d(GAGCAG_2T)$	Na^+	[46]
	G-T quartet	$d(G_3T)$	K^+	[19]
Pentads	A:(G:G:G:G) pentad	$d(G_3AG_2T_3G_3AT)$	Na^+	[47]
		$d(G_4TG_3AG_2AG_3T)$	K^+	[48]
Hexads	A:(G:G:G:G):A hexad	$d(G_2AG_2AN^a)$	Na^+	[49]
		$r(UG_2AG_2U)$	K^+	[50]
		$r(G_2AG_2U_4G_2AG_2)$	K^+	[51]
		$d(GCG_2AG_2AT)$	Na^+	[39]
	T:(G:G:G:G):T hexad	$d(GCG_2T_2G_2AT)$	Na^+	[39, 45]
Heptads	G-A heptad	$d(G_2A)_4$ and $d(G_2A)_8$	K^+	[52, 53]
Octads	G-C octad	$d(CG_2TG_2T)$	K^+ or NH_4^+	[18]
	G-U octad	$r(UG_4U)$	Sr^{2+}	[37]
		$d(^{Br}U^a)$-$r(GAG_2U)$	Ba^{2+} and Na^+	[32]
	A-U octad	$r(U)$-$d(^{Br}G^a)$-$r(AG_2U)$	Cobalt hexamine and Na^+	[33]

[a] ^{Br}U = 5-bromouracil; ^{Br}G = 8-bromoguanine; X = T, U, TT, UC, TC or A; Y = AT, AU, U or A; N = T or G

2.3 *Pentads, Hexads, Heptads, and Octads*

More than four nucleobases can be associated to form a planar cyclic array of H-bonds. The adenine or thymine in a single-residue propeller loop that bridges G-tetrad layers can form hydrogen bonds with one edge of the G-tetrad. We can observe pentad, hexad, heptad, or octad base alignments (Table 1) that contribute to G-quadruplex conformational diversity. Note that intramolecular or bimolecular quadruplexes may also involve base pairs or triads stacked on a terminal quartet; these arrangements are not common in tetramolecular complexes, but may occur through dimerization of tetramolecular species.

3 Stability and Association Kinetics of Tetramolecular Quadruplexes

3.1 Presentation

Tetramolecular quadruplexes are formed by short DNA or RNA sequences containing a single block of guanines. In this structure all strands are parallel, guanines are in the *anti* conformation with their attached sugars generally in C2′ *endo* (for DNA) and C3′ *endo* conformation (for RNA), and all groove dimension are identical [54]. Tetramolecular quadruplexes are kinetically inert: the association kinetics of the structure will be very slow, except at very high strand concentration, and once formed, it can be extremely stable. In the presence of potassium this structure can often resist high temperatures (95–100°C), so using lithium hydroxide (LiOH) is sometimes required to denature the complex.

The thermal denaturation of a G-quadruplex structure leads to a variation in absorbance that can be followed at specific wavelengths (such as 295 nm) [55]. The association process may also be followed over time at those same wavelengths at a constant temperature. These experiments enable one to determine thermal stability ($T_{1/2}$ = apparent melting temperature) and association kinetics (k_{on} = association rate constant) of the structures. Other techniques such as circular dichroism, fluorescence and NMR may also be used to follow thermal melting or association.

3.2 Association Process Analysis

It has been known for over 40 years that the renaturation of short guanine rich oligomers such as G_3 is slow [56]. In agreement with the paper published by Wyatt et al. [57], who used size exclusion chromatography to study this process, we [58] and others [59] found an experimental order of ≈4 for the association rate with respect to strand concentration. We used isothermal renaturation experiments to study the formation of the quadruplexes: starting from unfolded species, a time-dependent increase in absorbance at 295 nm and an opposite trend at 240 nm, with a time-dependent decrease in absorbance, are observed along G-quadruplex formation (Fig. 2a). The order of the reaction n may be experimentally estimated by analyzing the concentration dependency of the association process. Assuming that at $t = 0$, $\alpha = 1$, one can demonstrate that [57]

$$\alpha = \left(1 + C_0^{n-1}\,(n-1)\,k_{on}\,t\right)^{1/(1-n)}, \tag{1}$$

where C_0 is the initial strand concentration, α the fraction of unfolded-strand, and k_{on} the association rate constant. The constant k_{on} can be determined by fitting the association curves with this equation, considering that $n = 4$ or lies between 3 and

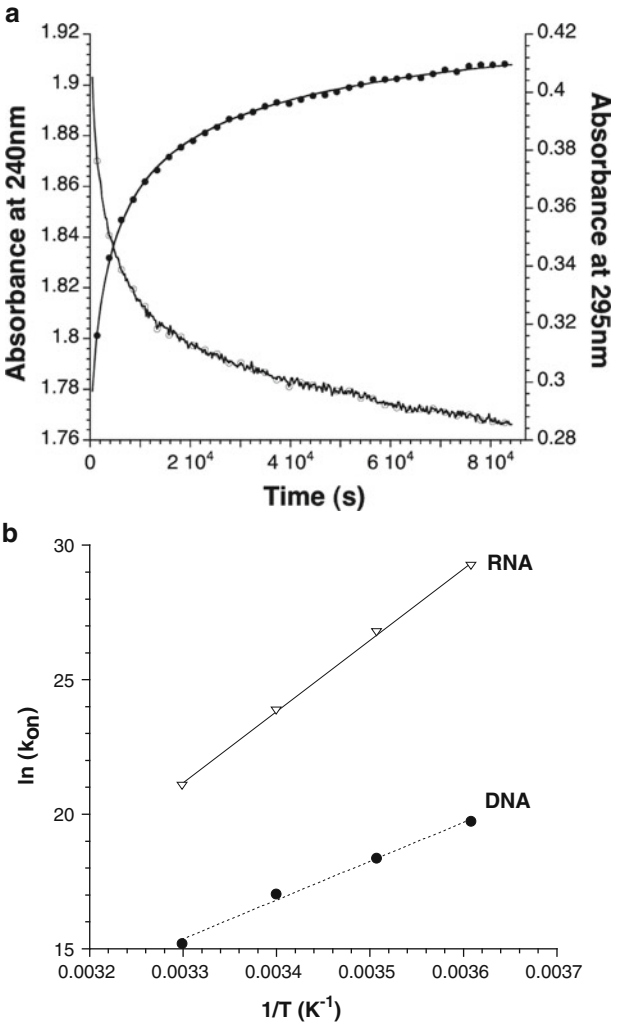


Fig. 2 *Tetramolecular quadruplex association*. (**a**) Example of an isothermal association experiment. Association at 4°C of d-TG_4T oligonucleotide at 50 μM strand concentration in a 0.1 M K^+ pH 7.2 buffer. Two different wavelengths are registered: 240 nm (*empty circles*, Y-scale shown on the *left*) and 295 nm (*black circle*, Y-scale shown on the *right*). (**b**) Arrhenius representation of the association rate ($\ln(k_{on})$) of d-TG_4T (DNA) and r-UG_4U (RNA) quadruplexes in Na^+. [Panel (**b**) reproduced from [142] with permission of the Royal Society of Chemistry]

4 [57–59] (Fig. 2b). Using NMR spectroscopy, Leroy and colleagues found a slightly lower experimental value for *n*, closer to 3 [60]. Thus, the exact value of the reaction order may depend on sequence and ionic conditions. Furthermore, NMR allows one to follow the association process in a more precise manner than absorbance or CD: it reveals the existence of "imperfect" tetramolecular

complexes, in which the strands are not aligned. These complexes may be long-lived, and their final conversion into the canonical quadruplex may take weeks or months [60, 61].

3.3 Thermal Stability Measurements

As seen before, the association kinetics of short parallel quadruplexes will be very slow at low strand concentration, and thermal denaturation of this structure is kinetically irreversible [58, 59, 62] when working in the nanomolar or micromolar concentration range, i.e., reassociation is negligible. The apparent melting temperature ($T_{1/2}$) determined from the denaturation curve depends on the temperature gradient since we do not obtain thermodynamic equilibrium at each temperature. In other words, variations in apparent T_m rather reflect the effects on dissociation (k_{off}) as a function of temperature than on the equilibrium constant. Little or no renaturation is found at low strand concentrations (Fig. 3a), while one may observe an extreme hysteresis at higher strand concentrations and slow rate of annealing [57, 58, 63].

One can take advantage of this quasi-irreversibility to determine dissociation rate constant (k_{off}) as a function of temperature (Fig. 3b), which represents an alternative approach to measure the lifetime of a pre-folded quadruplex using different unfolding temperature [57, 58, 63].

Analysis of CD melting spectra enables one to determine whether the thermal melting occurs as an all-or-nothing phenomenon or involves partially unfolded species. Petraccone et al. [59] demonstrated that the quadruplex to single-strands transition of d-TG_4T involved only two significantly populated spectral species, in full agreement with a simple dissociation pathway.

4 Stability and Association Kinetic Effectors for Tetramolecular Quadruplexes

Many parameters can influence the thermal stability and the association rate of the tetramolecular quadruplex; these parameters and a summary of their effects are presented in Table 2 and described below.

4.1 Effects of Temperature

Temperature can modulate both quadruplex association (k_{on}) and dissociation (k_{off}) rates [58, 63]. Because of the quasi-irreversible melting of tetramolecular

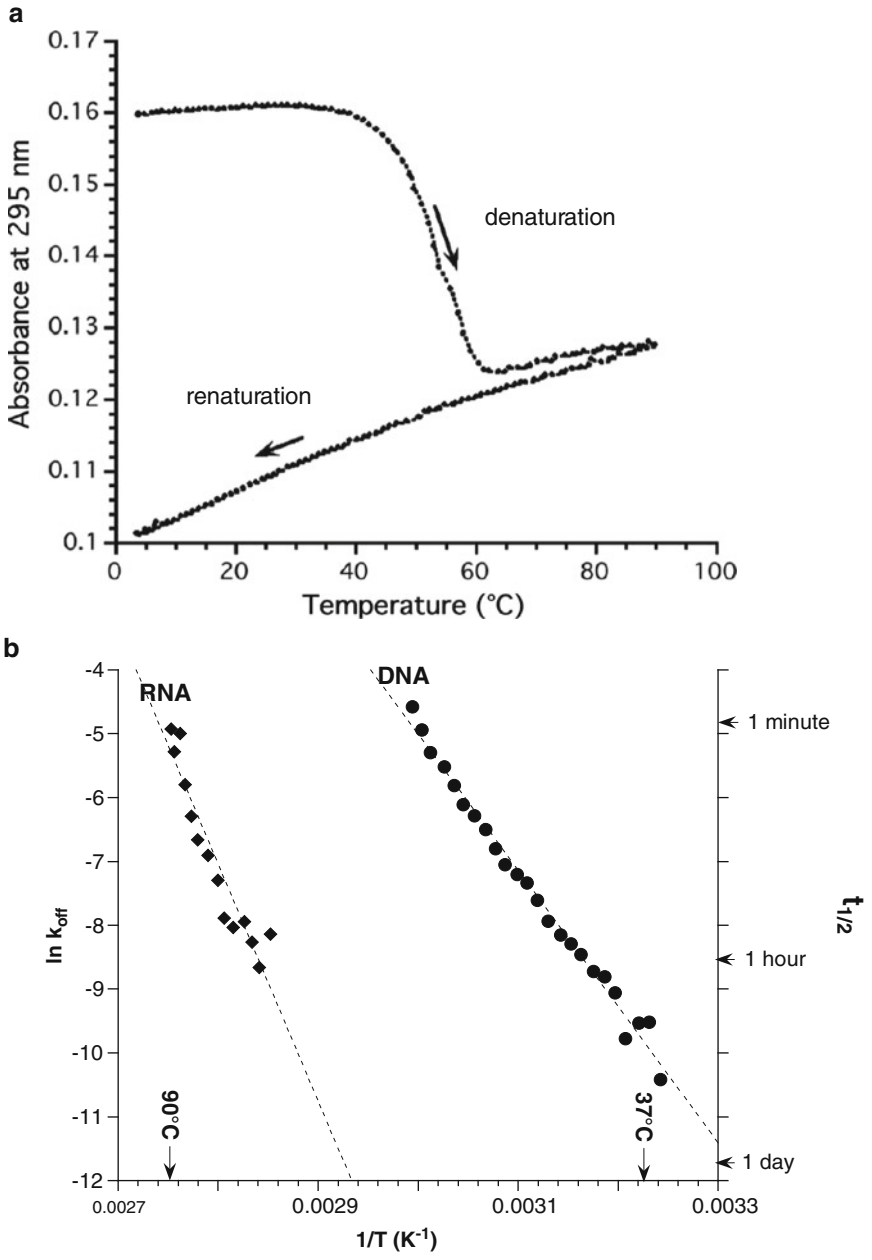


Fig. 3 *Tetramolecular quadruplex lifetime*. (**a**) Example of an irreversible melting curve of d-TG_4T in a 0.11 Na^+ pH 7.0 buffer recorded at 295 nm with a temperature gradient of ~0.2°C/min. *Arrows* indicate directions of temperature changes. The small difference observed at high temperature between the heating and cooling profiles is artefactual. [(**a**) reproduced from [58], by permission of Oxford University Press]. (**b**) Arrhenius representation of dissociation rate

Table 2 Effects of various parameters on stability and association rate of tetramolecular G-quadruplexes

Parameter	Association (k_{on})	Stability ($T_{1/2}$)
Decreased temperature	+	nd
Increased ionic strength	+	≈0
$Na^+ \rightarrow K^+$	+	+
Mg^{2+} addition	+	−[a]
Lower pH[b]	0	−
Increased oligonucleotide concentration	+[c]	≈0[c]
Longer G-tract	+	+
Longer non-G overhang (5′ and/or 3′ ends)	−[d]	+/−[d]
$^{5\prime}OH \rightarrow {}^{5\prime}P$	−	≈0

Note: "−" decrease, "+" increase, "0" little or no effect, "*nd*" not determined

[a] Magnesium concentration above 10 mM

[b] Below pH 5.5. No effect of pH in the 6.0–7.8 range

[c] k_{on} is obviously concentration independent, but apparent association rate strongly depends on concentration. $T_{1/2}$ does not depend on the strand concentration if reassociation is negligible (for low strand concentrations)

[d] Generally the addition of non-guanine bases stabilize the structure but sequence-dependent effects may be observed, depending on length, base composition (dT, dC or dA), and side (5′ or 3′) of the base insertion

quadruplexes, it is possible to study the association of these complexes at a low temperature where dissociation could be neglected – the lifetime of the complex is very long. Reciprocally, one can study at higher temperatures the dissociation process of preformed complexes; provided that strand concentration is low enough, reassociation can be neglected.

4.2 Effects of Ionic Strength

The stability and the association kinetics strongly depend on the nature of cation used ($K^+ > Na^+$). Cations, according to their ionic radii and their hydration energies, can stabilize more or less quadruplex structures (see Sect. 2.1). Potassium (ionic radius = 1.52 Å instead of 1.16 Å for Na^+) strongly stabilizes G-quadruplex structures leading to an increase in apparent melting temperature, which reflects a slower dissociation at a given temperature. It also increases the association rate constant by a factor of 20–50 as compared to other ionic conditions [58].

Monovalent cation concentration plays little if any role in tetramolecular quadruplex thermal stability. However, increasing ion concentration increases the association rate, as shown in Table 2. The addition of a divalent cation such as

Fig. 3 (continued) ($\ln(k_{off})$ on the *left* Y-scale) and lifetime (*right* Y-scale) of d-TG_4T (DNA) and r-UG_4U (RNA) quadruplexes in 0.11 M Na^+. [Panel (**b**) reproduced from [142] with permission of the Royal Society of Chemistry]

magnesium accelerates the association process in a concentration-dependent manner [58]. In contrast, increasing magnesium concentration above 10 mM leads to a decrease in the apparent melting temperature and hence induces an increase of the dissociation process. The trivalent cation Tb^{3+} stabilizes G-quadruplex structures in the low micromolar range; higher Tb^{3+} concentrations (above 10 μM) eventually destabilize the quadruplex [23].

4.3 Effects of pH

No variation of association or of thermal stability is observed in the 5.5–7.8 pH range [58]. However, pH below 5.5 accelerates both association and dissociation rates. The stabilizing effect on association is therefore compensated by a destabilizing effect on the denaturation process (a 4°C decrease in $T_{1/2}$).

4.4 Effects of Sequence and Concentration

The nucleic acid sequence can also modulate apparent melting temperature and association constant values [58]. Longer G-tracts can lead to a faster apparent association rate and a higher thermal stability, bearing in mind that kinetically trapped species with a lower number of quartets may delay the formation of the final "perfect" product. Each extra guanine induces an approximately tenfold increase in the association rate constant. In contrast, the addition of non-G-bases at the 5′ or 3′ end generally has a detrimental role in the association rate but can lead to an increase in the thermal stability of the DNA quadruplex. Besides, the replacement of a 5′ thymine by adenine has a beneficial effect on the association and the thermal stability. This reflects the importance of the nature of the 5′ terminal base on quadruplex formation.

In addition, increasing strand concentration can increase apparent association rate and so reduces reaction time, switching from a slow to a fast reaction [57]. However, $T_{1/2}$ does not depend on the strand concentration if the reassociation of the complexes is negligible (for low strand concentrations) [58].

4.5 Effects of the 5′-Phosphate Group

Our team demonstrated that the introduction of a 5′-terminal phosphate could affect quadruplex association, presumably due to the additional negative charges provided by the extra phosphate group. The 5′ phosphate group slowed the association rate but showed little effect on thermal stability [64].

5 Effects of Chaperone Molecules

Other than cations, several molecules (proteins, small ligands, and polymers) may act as molecular chaperones for tetramolecular quadruplex formation. Some of them will be presented below.

5.1 Effect of Proteins

Certain proteins can interact with tetramolecular G-quadruplex DNA and modulate the formation of the structures. Some of them cleave tetramolecular G-quadruplexes, such as GQN1 [65] and topoisomerase II [66], while other may act as chaperones which stabilize and/or promote G4 formation, such as topoisomerase I [67] or the nucleocapsid protein (NCp) of HIV-1 [68].

5.2 Effect of Quadruplex Ligands

Efforts are being made to identify small molecules that selectively bind to G-quadruplex structures as these ligands may have cellular effects interfering with the structures in vivo [4, 14–17]. Interestingly, ligands interacting with quadruplexes could also give clues to the dynamics of the structure. There are several possible interaction modes: end stacking, intercalation (between two quartets), and groove or loop interaction. Intercalation is generally considered unfavorable or unlikely, because it requires the transient unstacking/opening of two G-quartets and the release of a cation. Nevertheless, intercalation between a non-G-quartet and an external G-quartet has been reported [69] and solid evidence exists for "true" intercalation in the case of poly dG annealed in the absence of a stabilizing cation [70]. The most frequently observed mode of interaction is the end stacking mode, such as for TMPyP4, BRACO19, BSU6039, RHSP4, daunomycin, and naphthalene diimide [PDB entry: 2A5R, 3CE5, 1L1H, 1NZM, 100K, and 3SC8 (BMSG-SH3) respectively] [69, 71–75].

Furthermore, some of these small ligands may not only bind to and lock a preformed quadruplexes, leading to increase the lifetime of the structures (by lowering the dissociation constant, k_{off}), but also actively promote the formation of the complex and act as quadruplex chaperones (by increasing the association constant, k_{on}). Han et al. presented the first example of a small molecule, PIPER (a perylene derivative), that accelerates the assembly of di- and tetrameric G-quadruplexes formed by two tandem repeats of the human telomeric sequence (d-TTAG$_3$) [76]. We also investigated the effect of several well-characterized selective ligands on tetramolecular quadruplex formation. Among them, we demonstrated that 360A (bisquinolinium pyridine dicarboxamide compound) has the utmost capacity to act as a chaperone for tetramolecular complexes (d-TG$_3$T) by

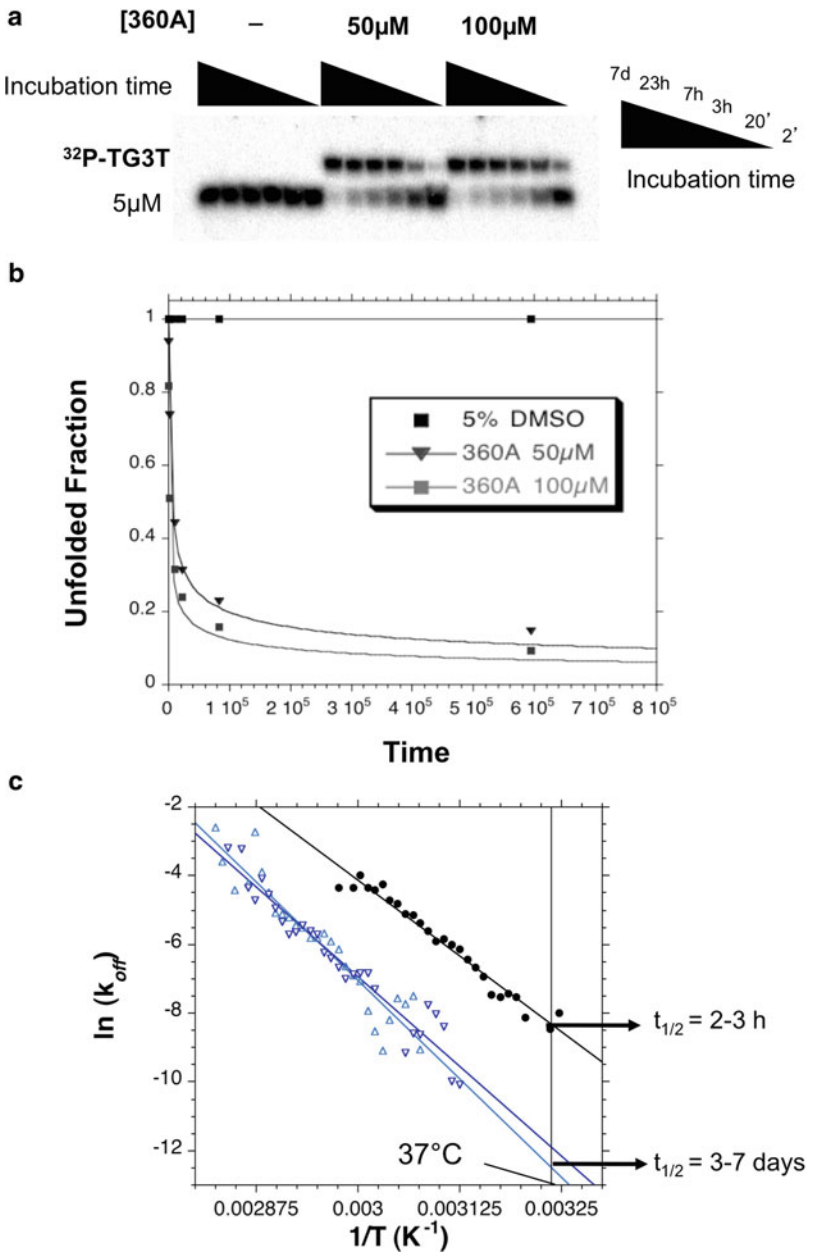


Fig. 4 *Small G-quadruplex ligand stabilizes tetramolecular quadruplex structures and accelerates their formation*. (**a**) Gel electrophoresis experiment: incubation of radiolabelled $^{32}PO_4$-$^{5'}TG_3T$ at 4°C at 5 μM strand concentration in absence or presence of a large excess of a specific quadruplex ligand, 360A, at 50 and 100 μM, for various amounts of time in a 10 mM lithium cacodylate buffer (pH 7.2) supplemented with 0.11 M K^+. Samples were loaded on 20%

increasing k_{on} (Fig. 4a, 4b) and by decreasing k_{off} (Fig. 4c) (i.e., an increase in the lifetime and thermal stability of the complex) [64]. This observation has implications for in vitro and in vivo applications of quadruplexes and should be taken into account when interpreting the cellular responses to these agents. Tetramolecular quadruplexes are well suited to study this chaperoning effect, as one can study and quantify G4 promoting activity under near-physiological conditions.

5.3 *Effect of Polymers*

Association rates of tetramolecular quadruplexes are extremely slow at micromolar strand concentration, likely owing to electrostatic repulsion among the four strands. Maruyama et al. have described how cationic comb-type copolymers comprised of more than 80% hydrophilic graft chains and less than 20 wt% cationic backbone (Fig. 5a) form water-soluble polyelectrolyte complexes with DNAs. The copolymers promote sequence-specific hybrid formation, such as formation of double-stranded DNA and triple-stranded DNA [78, 79], and they shield electrostatic repulsion of the DNA strands to stabilize DNA assembly [80–84]. Eventually they reported that this copolymer affects the kinetics of tetramolecular quadruplex formation by increasing the association rate (Fig. 5b) but also by accelerating the dissociation rate (Fig. 5c) [77, 85]. Moreover, the copolymer favors a strand exchange reaction between a preformed quadruplex and a single-strand to promote the reassembly into the thermodynamically most stable quadruplex structures [86, 87]. The cationic comb-type copolymer is therefore able to decrease the energy barrier for breakage or re-assembly of base pairings and convert meta-stable species into lowest energy species, as bona fide nucleic acid chaperones.

In the absence of cations, certain agents such as polyethylene glycol (PEG) can also stabilize the G-quadruplex structures [88] and influence G4 conformation: it favors the formation of parallel G-quadruplexes [89]. Other molecules such as polyamines (spermine or spermidine) [90] and triethylene tetraamine bind to G-quadruplex and

Fig. 4 (continued) acrylamide TBE 1X gel supplemented with 20 mM K^+. The tetramolecular G-quadruplex is too slow to associate in those conditions without the ligand. In presence of the ligand, a shift band reveals the formation of a tetramolecular quadruplex. (**b**) Isothermal association curves extrapolated from the gel. The data can be fitted by a fourth order model in presence of a large excess of 360A. The rate constant is increased by 2 and 21 million-fold, respectively, in presence of 50 and 100 μM of 360A (equivalent to an increase in strand concentration of 125- and 275-fold). (**c**) Arrhenius representation of dissociation rate [$\ln(k_{off})$] of $^{32}PO_4$-$^{5'}TG_3T$ quadruplexes in the absence (*black circle*) or the presence of 5 μM (*light blue triangle*) and 10 μM (*dark blue triangle*) 360A in 10 mM lithium cacodylate buffer (pH 7.2) supplemented with 0.11 M K^+. We observed that 360A decreases the dissociation rate of $^{32}PO_4$-$^{5'}TG_3T$ by a factor of ~33. [Panels (**a**) and (**b**) reproduced from [142] and [64] with permission of the Royal Society of Chemistry and of Oxford University Press

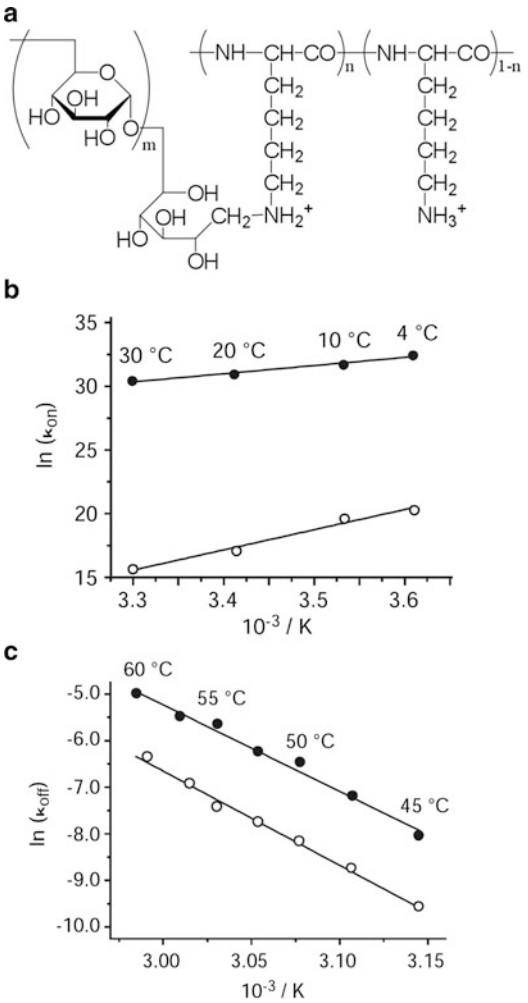


Fig. 5 *Cation comb-type polymer acts as a molecular chaperone for tetramolecular quadruplex formation*. (**a**) Structural formula of cationic comb-type polymer, poly(L-lysine)-*graft*-dextran. (**b**) Arrhenius representation of the association rate ($\ln(k_{on})$) of d-TG_4T in the absence (*empty circle*) or the presence (*black circle*) of PLL-*g*-Dex in 10 mM phosphate saline buffer (PBS) supplemented with 100 mM Na^+. (**c**) Arrhenius representation of dissociation rate ($\ln(k_{off})$) of d-TG_4T in the absence (*empty circle*) or the presence (*black circle*) of PLL-*g*-Dex in 10 mM PBS supplemented with 100 mM Na^+. (Figure reproduced from [77], with permission from Elsevier)

stabilize the structure [91]. It is important to mention that these compounds, generally described as crowding agents, also affect water activity, which makes more difficult the deconvolution of these two effects. Besides, Rosu et al. have also shown that the addition of methanol can significantly accelerate the cation-induced tetramolecular quadruplex assembly [61].

6 Effects of Chemical Modifications

In addition to their in vivo biological functions, quadruplexes may find use as pharmaceutically active agents [92]. A number of aptamers adopt a quadruplex fold, such as the thrombin-binding-aptamer (TBA) [93, 94] and the nucleolin binding-aptamer (AS1411) [95, 96]. While the previous examples are intra- or bi-molecular complexes, several tetramolecular quadruplexes were shown to act as specific inhibitors of human immunodeficiency virus type 1 (HIV-1) infection in vitro [97]. However, some of these sequences cannot be used in vivo because of their bioavailability problem. For this reason, the research on modified oligonucleotides can be useful to improve physical and biological properties of selected aptamers increasing biostability or resistance to nucleolytic enzymes [98]. The knowledge of modified nucleotide influence on thermal stability and kinetics of quadruplex formation is also crucial for their potential pharmacological applications.

6.1 RNA and 2′-Sugar Modified Analogs

RNAs (Fig. 6) containing short runs of Gs can form G-quadruplex structures. RNA quadruplexes, especially tetramolecular structures formed by r-UG$_n$U sequences, are much more stable than their DNA counterparts (d-TG$_n$T). This increased stability results from a much faster association (10^3-fold) and a slower dissociation (10^{-3}-fold) [58, 103]. Besides the contribution of the 2′OH group to the stability of the RNA quadruplex, another structural difference between the two structures is the organization of the uracil/thymine groups [36]. This stability can result in part from a stabilizing role of T → U base substitution [103] which may result from different characteristics of uracil and thymine quartets [20, 34, 36–38].

2′-Sugar modified analogs such as 2′-*O*-methyl ribonucleotide (O-RNA) [58] and locked nucleic acids (LNA) [104–106] (Fig. 6) form more stable tetramolecular quadruplexes than their natural DNA counterparts. From a kinetic point of view, this higher stability results from both a slower dissociation and a faster association. This difference might be explained by favoring the sugar in C3′-*endo* conformation which leads to a greater penalty in adopting a *syn* conformation for modified guanines [103, 106]. Data collected from our lab [58, 103] and other groups [106] allow oligomers to be ranked in the following order of $T_{1/2}$: DNA < O-RNA < RNA ≈ LNA. LNA quadruplexes, ribonucleotide analogs, present thermal stability similar to RNA complexes, but their association is second-order instead of third to fourth order for G4-DNA or RNA [106].

These results may have notable implications for a better understanding of the role of the 2′-position of the sugar moiety. The latter may participate in crucial interactions, which may be modulated by the possibility to form H bonds.

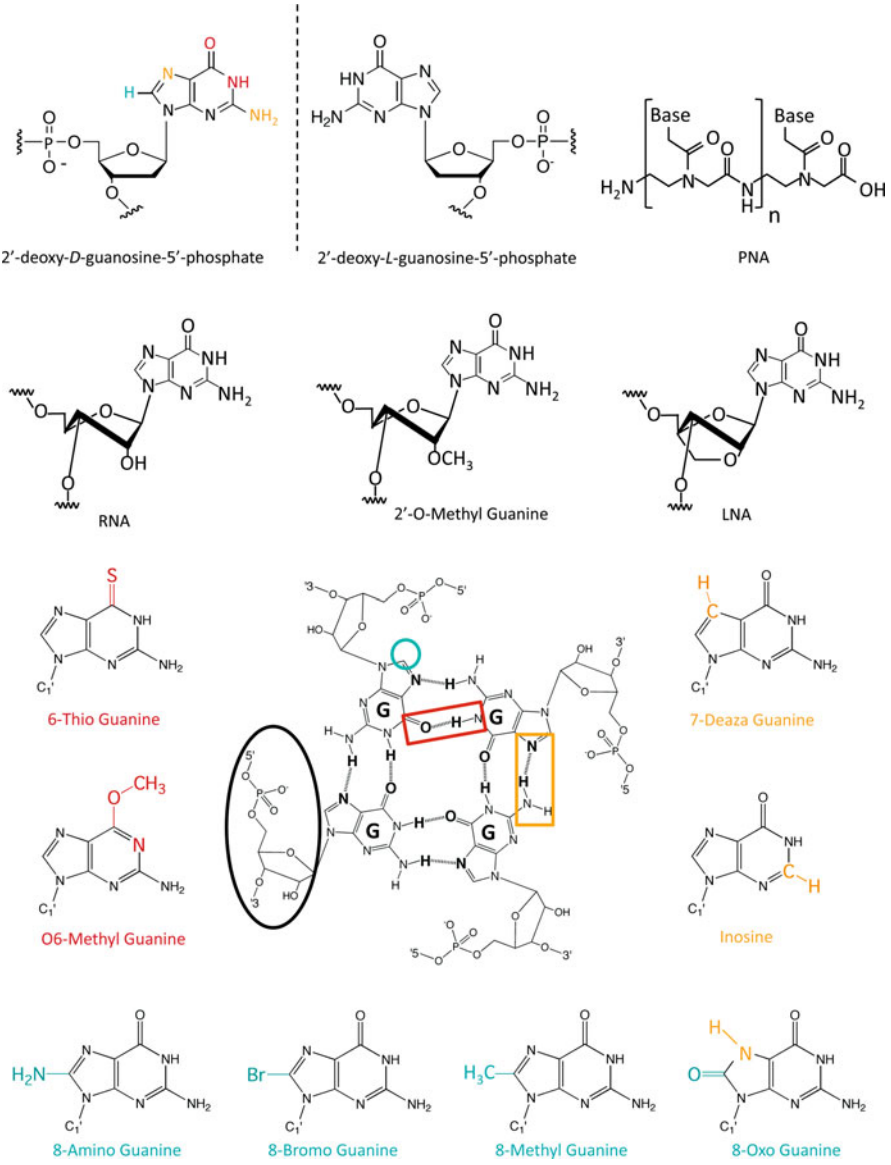


Fig. 6 *Backbone modifications and guanine analogs* [99–102]. In *black*: backbone modifications of guanine nucleotide: L-DNA, PNA, 2′-*O*-methylguanine and LNA. In *red*: base modifications disturbing the central ring of H-bonds from NH to C=O. In *orange*: base modifications perturbing the external ring of H-bonds from NH_2 to N. In *blue*: guanine modifications on the 8-position

Table 3 Effects of strand polarity inversion on structure, stability, and association rate of tetramolecular quadruplexes

Sequences	$T_{1/2}$ (°C)[a]	k_{on} (M^{-3} s^{-1})[a]	Stacking of quartets	Reference
$^{5'}$TGGGT$^{3'}$	47.5	8.3×10^7	a–a–a	[107]
$^{5'}$dS$^{3'}$-$^{3'}$GGGT$^{5'}$	42	2.7×10^7	a–a–a	
$^{5'}$T$^{3'}$-$^{3'}$GGGT$^{5'}$	47	9.1×10^7	a–a–a	
$^{5'}$TG$^{3'}$-$^{3'}$GGT$^{5'}$	51	t.s.	a–a–a	
$^{3'}$dS$^{5'}$-$^{5'}$GGGT$^{3'}$	41	t.s.	a–a–a	
$^{3'}$T$^{5'}$-$^{5'}$GGGT$^{3'}$	47.5	9.3×10^7	s–a–a	
$^{3'}$TG$^{5'}$-$^{5'}$GGT$^{3'}$	70	1.9×10^8	a–s–a	

Note: "*t.s.*" too slow, "a" *all-anti quartet*, "s" *all-syn quartet*, "*dS*" dSpacer (nucleoside without nucleobase)

[a] In 0.1 M K^+, 10 mM lithium cacodylate pH 7.2 buffer, with a temperature gradient of 0.4°C/min for the thermal denaturation experiment

6.2 *Further Backbone Modifications*

6.2.1 Strand Polarity Modifications

In 2009, Veronica Esposito et al. analyzed the effect of polarity site inversions (3′-3′ or 5′-5′) on the structure and the stability of tetramolecular G-quadruplexes (d-TG_3T) by NMR and circular dichroism [107] (Table 3). The introduction of inversion of polarity sites does not prevent G4 formation. In certain case this introduction may lead to the formation of all-*syn* quartets that may induce changes in CD spectra (Table 3) demonstrating that (1) tetramolecular G4 may involve guanines in *syn* and (2) all parallel G-quadruplexes do not always exhibit a "Type I" CD spectrum with a maximum at 260 nm. These modifications generally do not improve the stability or the association kinetics of the tetramolecular G4, except for the modified sequence d-TG$^{5'}$-$^{5'}$GGT. This modification appears to induce the formation of an all-*syn* quartet on the second position. This G4 structure displays higher apparent melting temperature and increases association rate constant by a factor of 2.

6.2.2 PNA

Peptide nucleic acid (PNA) is a DNA analog in which the entire sugar–phosphate backbone has been replaced with a pseudopeptide (Fig. 6), resulting in an achiral and neutral oligomer. Tetramolecular quadruplex formation with PNAs was first investigated for PNA:DNA hybrids [108]. Quadruplex invasion by PNAs may participate in the G4 formation and stabilization [109]. Indeed, tetramolecular quadruplexes resulting from PNA:DNA hybrids are more stable than their corresponding all-DNA structures [108]. Recently, it has been shown that the association kinetics of these PNA:DNA heteroquadruplexes may vary according

to the G4 conformations [110]: parallel G4-DNA structures in which all guanines are in the *anti* conformation enable fastest PNA hybridization kinetics.

It is also possible to form tetramolecular quadruplex with only PNA strands, but G4-PNA is less stable than G4-DNA [111]. This may be due to the flexibility of the PNA backbone that probably has an unfavorable impact on the energetics of quadruplex formation. In contrast to the canonical G4-DNA or RNA, G4-PNA formed by tggg (lower case letter refers to PNAs) is not much stabilized, or is even destabilized by cations, as a result of an uncharged PNA backbone [111].

It is also possible to design a chimeric strand bearing DNA and PNA residues such as TGGGt. The thermal stability of the chimera formed by TGGGt is quite similar to that observed for the DNA equivalent (TGGGT) [112]. However, Petraccone et al. observed a higher thermal stability (4–7°C) for chimeric quadruplexes from tGGGGT and TGGGGt as compared to the corresponding TG_4T DNA strand [59].

6.2.3 Mirror Image Quadruplexes

Nucleic acids are chiral molecules and thus contain enantiomers. L-Nucleic acids (Fig. 6) are the mirror images of natural D-DNA or D-RNA. L-Nucleic acids have found a number of interesting applications in biotechnology [113, 114] (as aptamers or molecular beacons) and nanotechnology [115]. One of the advantages of L-nucleic acids is their nuclease resistance, allowing two *spiegelmers* (from "Spiegel" meaning mirror in German) to enter clinical trials [86, 116]. L-DNA forms duplexes with identical physical characteristics (solubility and stability) except for chirality, leading to left-handed double-helices [87]. However, the heterochiral duplex resulting from L- and D-strand hybrids is not stable [117]. G-quadruplex formation with L-DNA was first investigated by Urata et al. in 2002 [118]. They proposed that L-dG rich sequences are also able to form G-quadruplexes. L-d($GTG_2TG_3TG_3TG_3T$), the L-enantiomer of Zintevir, displays a similar G4 structure and thermal stability in comparison with unmodified sequence, except for chirality. Furthermore, this Zintevir analog showed a comparable or event higher anti-HIV-1 activity than that of Zintevir. This higher anti-HIV-1 activity is probably due to the nuclease resistance property of the L-enantiomer. Recently, our team investigated the potential of a short TG_4T G-rich L-DNA strand to form a tetramolecular quadruplex [99], using biophysical and biochemical approaches. We demonstrated that this left-handed quadruplex has the same properties (similar thermal stability and association kinetics) as a D-DNA strand of identical sequence, besides an opposite chirality resulted in an inverted circular dichroism spectrum. Hybrid G-quadruplexes do not form when mixing L- and D-strands together: enantiomeric pure G-quadruplexes are formed. In parallel the team of Pr. Galeone analyzed the formation of L-quadruplexes with chimeric strands bearing D- and L-DNA residues such as TG**GG**GT (bold letters refer to L-DNAs). They showed that the position of L-residues in the sequence affects the chirality and the thermal stability of the quadruplexes: the 3′ end and the core of the G-run are more important than the 5′

end in determining quadruplex helicity (right- or left-handed). Some modified sequences such as TG**GG**GT can lead to the formation of an all-*syn* G-quartet.

6.3 Guanine Base Modifications

Quadruplexes may incorporate unusual quartets with canonical nucleobases (see Sect. 2.2) or guanine analogs. These modifications may affect the DNA quadruplex structure, its apparent melting temperature ($T_{1/2}$), and/or its association rate (k_{on}). During the last decade, several teams [119–123], including ours [100, 101, 124], have been interested in the effect of guanine analogs on G4 formation. The analogs presented here are classified according to their impact on H-bonds between guanine bases of the same quartet (Fig. 6). 6-Thioguanine and O^6-methylguanine (in red) disturb the central ring of H-bonds between O^6 of carbonyl group and H^1 of nitrogen from two neighboring guanine bases. Inosine and 7-deazaguanine (in orange) disturb the internal ring of H-bonds between N^7 and H^2 of amine. 8-Oxoguanine, 8-aminoguanine, 8-bromoguanine, and 8-methylguanine (in blue) replace H^8 of canonical guanine with bulkier and/or charged groups. We analyzed the effect of these modifications on tetramolecular quadruplex formation ($T_{1/2}$ and k_{on}) and obtained modification- and position-dependent effects. The formation of a tetramolecular quadruplex containing a single modified base is possible but most modifications have an unfavorable contribution to the association and dissociation rates of the G4 complex [100]. The inner part of a G-quartet (i.e., the central ring of H-bonds) is crucial for its stability; disturbing this part not only leads to the loss of one H-bond, but may also alter coordination of the central cation with the carbonyl groups (in the case of 6-thioG and 6-methylG). Removal of the external ring of H-bonds generally leads to a moderate decrease in the association rate (observed in the case of inosine). 7-Deazaguanine not only prevents this external H-bond ring; it also perturbs the geometry/planarity of the quartet because of steric clashes, leading to a more severe effect on both association and dissociation rates.

Nevertheless, not all modifications are detrimental, such as 8-aminoG, 8-bromoG, and 8-methylG. These guanine analogs generally favor G-quadruplex formation since they do not disturb the cyclic array of H-bonds of the G-quartet. However, position-dependent effects are found. Our team observed that 8-aminoguanine accelerates tetramolecular quadruplex formation when inserted at the 5′ end (63-fold faster than unmodified sequence) but has a detrimental effect on the thermal stability of the quadruplexes [101]. In contrast, the insertion of 8-aminoG at central positions (d-TGRGGT where R = 8-aminoG) leads to an increase in the apparent melting temperature. We also showed that 8-bromoguanine and 8-methylguanine stabilize and accelerate quadruplex formation by a factor of 10 [100] and 15 [125], respectively, when present at the 5′ end. This insertion position of the modifications favors the formation of an all-*syn* quartet at the 5′ end [120] that may suggest the implication of *syn* G at the 5′ end in the nucleation process. This formation of all-*syn* quartet at the 5′ end also stabilizes the complexes.

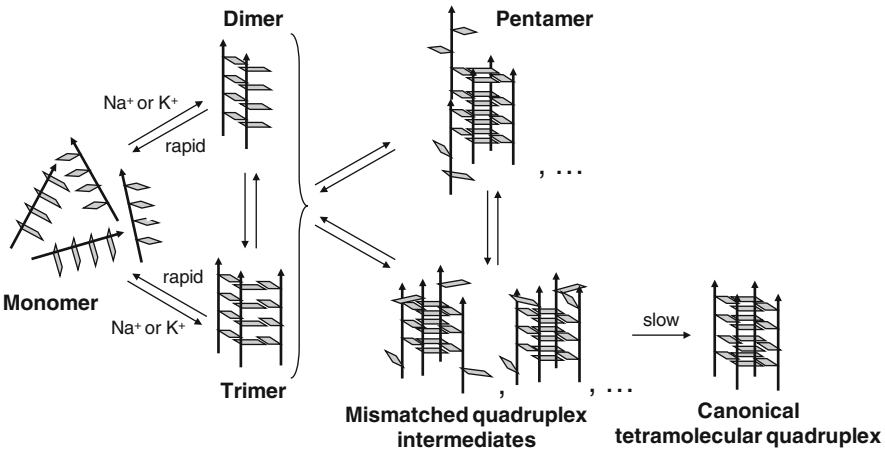


Fig. 7 Potential folding pathway of tetramolecular quadruplexes as reported in [60, 61, 126]

Indeed, molecular modeling suggested that the stacking of an all-*syn* quartet above an all-*anti* quartet was more stable than the reverse situation (Dr. A.T. Phan, personal communication).

7 Potential Folding Pathway of Tetramolecular Quadruplexes

In 2003, results from molecular dynamic simulations published by Richard Stefl et al. suggested that the association pathway of tetramolecular quadruplex might have several stable intermediate states [126]. They described a stepwise strand addition model: (1) monomers, (2) two-stranded assembly (dimers), (3) the addition of third strand (trimers), and eventually (4) the slower formation of four-stranded structure (tetramers) (Fig. 7). They also suggested the formation of four-stranded intermediates containing several slipped strands. These results were supported by the NMR study of Leroy's team in 2007 [60] and more recently by the mass spectrometry study of Rosu et al. [61]. Leroy and co-worker showed that the formation of mismatched quadruplex (containing slipped strands) is faster than the canonical species even if the latter displays a higher stability. As mismatched species have relatively long lifetimes, their formation may hinder the association of fully matched canonical tetramolecular quadruplex. Indeed, Rosu et al. detected the presence of these four-stranded intermediate species by electrospray mass spectrometry thanks to their long lifetime. Furthermore, they observed the formation of five-stranded species (pentamers). They also showed that the rearrangement of slipped strands, the suppression of the fifth strand slippage, and the re-incorporation of the last cation into canonical structure can be extremely slow; the process may last more than 4 months at 4°C. We also observed this phenomenon with modified sequences such as d-TGMGGT and d-TGGGMT (M = 8-methylguanine) [124]:

circular dichroism signals registered after 2-day incubation is different from that measured after 1 month incubation at 4°C. While the existence of intermediate dimer- and trimer-species is experimentally verified, their nature remains elusive. Figure 7 depicts these intermediates as duplexes or triplexes, but there is little direct evidence for such structures; we propose an alternative model in which these dimers and trimers already involve the formation of one or more quartets – in that case, two or more guanines must come from the same strand.

8 Biotechnology and Nanotechnology Applications

Besides their biological importance, the self-assembly, stability, and stiffness properties of G-quadruplex structures make them particularly interesting for nanotechnology and biotechnology applications. Efforts in using tetramolecular quadruplexes to build supramolecular structures are described below.

8.1 DNA Nanostructures: G-Wires, Frayed Wires, and "Synapsable" DNA

G-wires are obtained by the stacking of tetramolecular G-quadruplex formed by telomeric sequence of some ciliates (d-$G_4T_2G_4$) [127] or by poly-dG sequences [128]. These nanowires have high conductivity potential which results from electronic transport created by central cation coordination [129]. This conductivity of the G-wires is of potential interest for microelectronic applications. Furthermore, it is possible to form a G-wire with single-stranded overhangs that enables the assembly of complex structures for nanofabrication purposes [130].

MacGregor and co-workers first identified the structure of frayed wires formed by d-$A_{15}G_{15}$ or d-$T_{15}G_{15}$ sequences [131]. These frayed wires are distinguished from G-wires by their flexible A_{15}/T_{15} tails that radiate from a guanine core. The poly-dA/dT tails enable the attachment of objects such as enzymes [132].

"Synapsable" DNA was described by Sen and co-workers in 1996 [133]. It results from the interaction between two DNA duplexes containing mismatched regions of repeating G–G base pairs. The repeating G–G mismatches facilitate the formation of G-quartets that leads to the dimerization of the duplex. In contrast with all tetramolecular complexes described so far, the tetramolecular G-quadruplex formed within the "synaptic domain" is an antiparallel structure. As for parallel tetramolecular quadruplexes, this structure has slow association kinetics but, once formed, it is extremely stable. Their formation depends on ionic strength which follows the trend typically observed for classical G-quartet stabilization (see Sect. 2.1) [134]. The formation of "synaptic" DNA is possible between identical duplexes (homodimers) or between two different double-helices (heterodimers).

This original structure provides a means for the self-recognition and supramolecular assembly; it can also be used as a tool for investigating certain protein-DNA interaction complexes such as topoisomerase I or GQN1, which bind to tetramolecular quadruplexes.

8.2 Quadruplex as Ion Channels

As already mentioned, cation movement through G4 structures makes them a potential ion channel. Indeed, it is possible to build a synthetic transmembrane ion channel from the stacking of lipophilic G-quartets [135–137]. This lipophilic G4 structure allows efficient cation transport through a lipid bilayer.

8.3 Nanoparticles Connected via Tetramolecular Quadruplexes

Kotlyar's team showed that the parallel tetramolecular G-quadruplex containing phosphorothioate residues at both strand ends can anchor on the particle surface and enable silver [138] or gold [139] nanoparticle assembly, leading to the formation of flower-shape structures. These nanoparticles connected via G-quadruplex structures allow stable assembly without inducing aggregation even at high salt concentration (0.2 M).

9 Conclusion

G-quadruplexes can be formed not only by regular G-quartets but also by non-canonical planar base associations. This contributes to the structural polymorphism of G-quadruplexes and especially the tetramolecular complexes, which are considered as the simplest model for quadruplex formation study. Of course, the thermal stability and the association kinetics of tetramolecular quadruplexes depend strongly on sequence, but also on experimental conditions, such as the nature and concentration of cations, temperature, and pH (see Sect. 4). Natural and synthetic molecules (protein, ligands, and polymers) can also affect these parameters (stability and association rates); they may act as molecular chaperones for tetramolecular quadruplex assembly (see Sect. 5). Besides, efforts are being made to study the effect of different chemical modifications on quadruplex formation in order to improve the physical and biological properties of these structures. Such modifications may increase biostability or nuclease resistance crucial for therapeutic application (see Sect. 6), for example mirror image nucleic acids. Some chemical modifications can be useful for nanotechnology or biotechnology applications. Sugimoto's team presented the formation of modified G-wires with 2,2′-bipyridine

units instead of the thymines of d-$G_4T_4G_4$. This modification allows one to control G-wire switching by external signals that is important for the construction of molecular electronic technologies [140, 141]. The use of lipophilic guanosines also enhances the rigidity of the G-quadruplex structures and then allows efficient cation transport through the mimic membrane [136, 137].

The study of tetramolecular quadruplex stability and assembly allows one to understand better the rules governing their formation and to suggest a potential folding pathway of the structure (see Sect. 7). The latter reveals the long lifetime of some intermediate species that could explain the slow formation of these four-stranded structures. Despite important progress in our understanding of structural properties, stability, and kinetics of tetramolecular quadruplexes, considerable work remains to be done for bi- and unimolecular conformations.

Acknowledgments This work has been supported by ANR (QuantADN, F-DNA and G4-Toolbox), *Conseil Régional d'Aquitaine*, ARC and FRM. We would like to thank all our collaborators, past and present, for helpful discussions.

References

1. Gellert M, Lipsett MN, Davies DR (1962) Helix formation by guanylic acid. Proc Natl Acad Sci USA 48:2013–2018
2. Huppert JL, Balasubramanian S (2005) Prevalence of quadruplexes in the human genome. Nucleic Acids Res 33(9):2908–2916
3. Lopes J et al (2011) G-quadruplex-induced instability during leading-strand replication. EMBO J 30(19):4033–4046
4. Rodriguez R et al (2012) Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. Nat Chem Biol 8(3):301–310
5. Piazza A et al (2010) Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in *Saccharomyces cerevisiae*. Nucleic Acids Res 38(13):4337–4348
6. Ribeyre C et al (2009) The yeast Pif1 helicase prevents genomic instability caused by G-quadruplex-forming CEB1 sequences in vivo. PLoS Genet 5(5):e1000475
7. De Cian A et al (2008) Plasmodium telomeric sequences: structure, stability and quadruplex targeting by small compounds. Chembiochem 9(16):2730–2739
8. Sarkies P et al (2012) FANCJ coordinates two pathways that maintain epigenetic stability at G-quadruplex DNA. Nucleic Acids Res 40(4):1485–1498
9. Paeschke K, Capra JA, Zakian VA (2011) DNA replication through G-quadruplex motifs is promoted by the *Saccharomyces cerevisiae* Pif1 DNA helicase. Cell 145(5):678–691
10. Gomez D et al (2010) A G-quadruplex structure within the 5′-UTR of TRF2 mRNA represses translation in human cells. Nucleic Acids Res 38(20):7187–7198
11. Wu Y, Shin-ya K, Brosh RM Jr (2008) FANCJ helicase defective in Fanconia anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability. Mol Cell Biol 28(12):4116–4128
12. Qin Y, Hurley LH (2008) Structures, folding patterns, and functions of intramolecular DNA G-quadruplexes found in eukaryotic promoter regions. Biochimie 90(8):1149–1171
13. Kumari S et al (2007) An RNA G-quadruplex in the 5′ UTR of the NRAS proto-oncogene modulates translation. Nat Chem Biol 3(4):218–221
14. Gonzalez V, Hurley LH (2010) The c-MYC NHE III(1): function and regulation. Annu Rev Pharmacol Toxicol 50:111–129

15. Riou JF et al (2002) Cell senescence and telomere shortening induced by a new series of specific G-quadruplex DNA ligands. Proc Natl Acad Sci USA 99(5):2672–2677
16. Drygin D et al (2009) Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis. Cancer Res 69(19):7653–7661
17. Cookson JC et al (2005) Pharmacodynamics of the G-quadruplex-stabilizing telomerase inhibitor 3,11-difluoro-6,8,13-trimethyl-8H-quino[4,3,2-kl]acridinium methosulfate (RHPS4) in vitro: activity in human tumor cells correlates with telomere length and can be enhanced, or antagonized, with cytotoxic agents. Mol Pharmacol 68(6):1551–1558
18. Borbone N et al (2011) d(CGGTGGT) forms an octameric parallel G-quadruplex via stacking of unusual G(:C):G(:C):G(:C):G(:C) octads. Nucleic Acids Res 39(17):7848–7857
19. Krishnan-Ghosh Y, Liu D, Balasubramanian S (2004) Formation of an interlocked quadruplex dimer by d(GGGT). J Am Chem Soc 126(35):11009–11016
20. Sket P, Plavec J (2010) Tetramolecular DNA quadruplexes in solution: insights into structural diversity and cation movement. J Am Chem Soc 132(36):12724–12732
21. Smith FW, Lau FW, Feigon J (1994) d(G3T4G3) forms an asymmetric diagonally looped dimeric quadruplex with guanosine 5′-syn-syn-anti and 5′-syn-anti-anti N-glycosidic conformations. Proc Natl Acad Sci USA 91(22):10546–10550
22. Smith FW, Feigon J (1992) Quadruplex structure of oxytricha telomeric DNA oligonucleotides. Nature 356(6365):164–168
23. Galezowska E, Gluszynska A, Juskowiak B (2007) Luminescence study of G-quadruplex formation in the presence of Tb3+ ion. J Inorg Biochem 101(4):678–685
24. Worlinsky JL, Basu S (2009) Detection of quadruplex DNA by luminescence enhancement of lanthanide ions and energy transfer from lanthanide chelates. J Phys Chem B 113(4):865–868
25. Wong A, Wu G (2003) Selective binding of monovalent cations to the stacking G-quartet structure formed by guanosine 5′-monophosphate: a solid-state NMR study. J Am Chem Soc 125(45):13895–13905
26. Venczel EA, Sen D (1993) Parallel and antiparallel G-DNA structures from a complex telomeric sequence. Biochemistry 32(24):6220–6228
27. Sket P, Plavec J (2007) Not all G-quadruplexes exhibit ion-channel-like properties: NMR study of ammonium ion (non)movement within the d(G(3)T(4)G(4))(2) quadruplex. J Am Chem Soc 129(28):8794–8800
28. Hud NV et al (1999) Binding sites and dynamics of ammonium ions in a telomere repeat DNA quadruplex. J Mol Biol 285(1):233–243
29. Podbevsek P, Sket P, Plavec J (2008) Stacking and not solely topology of T3 loops controls rigidity and ammonium ion movement within d(G4T3G4)2 G-quadruplex. J Am Chem Soc 130(43):14287–14293
30. Podbevsek P, Hud NV, Plavec J (2007) NMR evaluation of ammonium ion movement within a unimolecular G-quadruplex in solution. Nucleic Acids Res 35(8):2554–2563
31. Patel PK, Koti AS, Hosur RV (1999) NMR studies on truncated sequences of human telomeric DNA: observation of a novel A-tetrad. Nucleic Acids Res 27(19):3836–3843
32. Pan B et al (2003) Crystal structure of an RNA purine-rich tetraplex containing adenine tetrads: implications for specific binding in RNA tetraplexes. Structure 11(7):815–823
33. Pan B et al (2003) An eight-stranded helical fragment in RNA crystal structure: implications for tetraplex interaction. Structure 11(7):825–831
34. Patel PK, Hosur RV (1999) NMR observation of T-tetrads in a parallel stranded DNA quadruplex formed by *Saccharomyces cerevisiae* telomere repeats. Nucleic Acids Res 27(12):2457–2464
35. Patel PK, Bhavesh NS, Hosur RV (2000) NMR observation of a novel C-tetrad in the structure of the SV40 repeat sequence GGGCGG. Biochem Biophys Res Commun 270(3):967–971
36. Cheong C, Moore PB (1992) Solution structure of an unusually stable RNA tetraplex containing G- and U-quartet structures. Biochemistry 31(36):8406–8414
37. Deng J, Xiong Y, Sundaralingam M (2001) X-Ray analysis of an RNA tetraplex (UGGGGU) (4) with divalent Sr(2+) ions at subatomic resolution (0.61 A). Proc Natl Acad Sci USA 98(24):13665–13670

38. Pan B, Shi K, Sundaralingam M (2006) Base-tetrad swapping results in dimerization of RNA quadruplexes: implications for formation of the i-motif RNA octaplex. Proc Natl Acad Sci USA 103(9):3130–3134
39. Mergny JL et al (2006) Kinetics of double-chain reversals bridging contiguous quartets in tetramolecular quadruplexes. Nucleic Acids Res 34(8):2386–2397
40. Kettani A, Kumar RA, Patel DJ (1995) Solution structure of a DNA quadruplex containing the fragile X syndrome triplet repeat. J Mol Biol 254(4):638–656
41. Kettani A et al (1998) Solution structure of a Na cation stabilized DNA quadruplex containing G.G.G.G and G.C.G.C tetrads formed by G-G-G-C repeats observed in adeno-associated viral DNA. J Mol Biol 282(3):619–636
42. Bouaziz S, Kettani A, Patel DJ (1998) A K cation-induced conformational switch within a loop spanning segment of a DNA quadruplex containing G-G-G-C repeats. J Mol Biol 282 (3):637–652
43. Zavasnik J, Podbevsek P, Plavec J (2011) Observation of water molecules within the bimolecular d(GCTGC)G-quadruplex. Biochemistry 50(19):4155–4161
44. Lim KW et al (2009) Sequence variant (CTAGGG)n in the human telomere favors a G-quadruplex structure containing a G.C.G.C tetrad. Nucleic Acids Res 37(18):6239–6248
45. Webba da Silva M (2005) Experimental demonstration of T:(G:G:G:G):T hexad and T:A:A:T tetrad alignments within a DNA quadruplex stem. Biochemistry 44(10):3754–3764
46. Zhang N et al (2001) Dimeric DNA quadruplex containing major groove-aligned A-T-A-T and G-C-G-C tetrads stabilized by inter-subunit Watson-Crick A-T and G-C pairs. J Mol Biol 312(5):1073–1088
47. Zhang N et al (2001) V-shaped scaffold: a new architectural motif identified in an A x (G x G x G x G) pentad-containing dimeric DNA quadruplex involving stacked G(*anti*) x G(*anti*) x G (*anti*) x G(*syn*) tetrads. J Mol Biol 311(5):1063–1079
48. Phan AT et al (2005) An interlocked dimeric parallel-stranded DNA quadruplex: a potent inhibitor of HIV-1 integrase. Proc Natl Acad Sci USA 102(3):634–639
49. Kettani A et al (2000) A dimeric DNA interface stabilized by stacked A.(G.G.G.G).A hexads and coordinated monovalent cations. J Mol Biol 297(3):627–644
50. Lipay JM, Mihailescu MR (2009) NMR spectroscopy and kinetic studies of the quadruplex forming RNA r(UGGAGGU). Mol Biosyst 5(11):1347–1355
51. Liu H et al (2002) A dimeric RNA quadruplex architecture comprised of two G:G(:A):G:G (:A) hexads, G:G:G:G tetrads and UUUU loops. J Mol Biol 322(5):955–970
52. Matsugami A et al (2001) An intramolecular quadruplex of (GGA)(4) triplet repeat DNA with a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad, and its dimeric interaction. J Mol Biol 313(2):255–269
53. Matsugami A et al (2003) Intramolecular higher order packing of parallel quadruplexes comprising a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad of GGA triplet repeat DNA. J Biol Chem 278(30):28147–28153
54. Phillips K et al (1997) The crystal structure of a parallel-stranded guanine tetraplex at 0.95 A resolution. J Mol Biol 273(1):171–182
55. Mergny JL, Phan AT, Lacroix L (1998) Following G-quartet formation by UV-spectroscopy. FEBS Lett 435(1):74–78
56. Ralph RK, Connors WJ, Khorana HG (1962) Secondary structure and aggregation in deoxyguanosine oligonucleotides. J Am Chem Soc 84(11):2265–2266
57. Wyatt JR, Davis PW, Freier SM (1996) Kinetics of G-quartet-mediated tetramer formation. Biochemistry 35(24):8002–8008
58. Mergny JL et al (2005) Kinetics of tetramolecular quadruplexes. Nucleic Acids Res 33(1):81–94
59. Petraccone L et al (2005) Thermodynamics and kinetics of PNA-DNA quadruplex-forming chimeras. J Am Chem Soc 127(46):16215–16223
60. Bardin C, Leroy JL (2008) The formation pathway of tetramolecular G-quadruplexes. Nucleic Acids Res 36(2):477–488
61. Rosu F et al (2010) Tetramolecular G-quadruplex formation pathways studied by electrospray mass spectrometry. Nucleic Acids Res 38(15):5217–5225

62. Mergny JL et al (2005) Thermal difference spectra: a specific signature for nucleic acid structures. Nucleic Acids Res 33(16):e138
63. Merkina EE, Fox KR (2005) Kinetic stability of intermolecular DNA quadruplexes. Biophys J 89(1):365–373
64. De Cian A, Mergny JL (2007) Quadruplex ligands may act as molecular chaperones for tetramolecular quadruplex formation. Nucleic Acids Res 35(8):2483–2493
65. Sun H, Yabuki A, Maizels N (2001) A human nuclease specific for G4 DNA. Proc Natl Acad Sci USA 98(22):12444–12449
66. Chung IK et al (1992) Eukaryotic topoisomerase II cleavage of parallel stranded DNA tetraplexes. Nucleic Acids Res 20(8):1973–1977
67. Arimondo PB et al (2000) Interaction of human DNA topoisomerase I with G-quartet structures. Nucleic Acids Res 28(24):4832–4838
68. Lyonnais S et al (2003) G-quartets direct assembly of HIV-1 nucleocapsid protein along single-stranded DNA. Nucleic Acids Res 31(19):5754–5763
69. Gavathiotis E et al (2003) Drug recognition and stabilisation of the parallel-stranded DNA quadruplex d(TTAGGGT)4 containing the human telomeric repeat. J Mol Biol 334(1):25–36
70. Lubitz I, Borovok N, Kotlyar A (2007) Interaction of monomolecular G4-DNA nanowires with TMPyP: evidence for intercalation. Biochemistry 46(45):12925–12929
71. Collie GW et al (2012) Structural basis for telomeric G-quadruplex targeting by naphthalene diimide ligands. J Am Chem Soc 134(5):2723–2731
72. Haider SM, Parkinson GN, Neidle S (2003) Structure of a G-quadruplex-ligand complex. J Mol Biol 326(1):117–125
73. Clark GR et al (2003) Structure of the first parallel DNA quadruplex-drug complex. J Am Chem Soc 125(14):4066–4067
74. Campbell NH et al (2008) Structural basis of DNA quadruplex recognition by an acridine drug. J Am Chem Soc 130(21):6722–6724
75. Phan AT et al (2005) Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. Nat Chem Biol 1(3):167–173
76. Han H, Cliff CL, Hurley LH (1999) Accelerated assembly of G-quadruplex structures by a small molecule. Biochemistry 38(22):6981–6986
77. Moriyama R et al (2011) DNA assembly and re-assembly activated by cationic comb-type copolymer. Biomaterials 32(9):2351–2358
78. Maruyama A et al (1997) Comb-type polycations effectively stabilize DNA triplex. Bioconjug Chem 8(1):3–6
79. Maruyama A et al (1998) Characterization of interpolyelectrolyte complexes between double-stranded DNA and polylysine comb-type copolymers having hydrophilic side chains. Bioconjug Chem 9(2):292–299
80. Torigoe H et al (2009) Synergistic stabilization of nucleic acid assembly by 2′-O,4′-C-methylene-bridged nucleic acid modification and additions of comb-type cationic copolymers. Biochemistry 48(15):3545–3553
81. Kim WJ et al (2001) Comb-type cationic copolymer expedites DNA strand exchange while stabilizing DNA duplex. Chemistry 7(1):176–180
82. Torigoe H et al (1999) Poly(L-lysine)-graft-dextran copolymer promotes pyrimidine motif triplex DNA formation at physiological pH. Thermodynamic and kinetic studies. J Biol Chem 274(10):6161–6167
83. Kim WJ, Akaike T, Maruyama A (2002) DNA strand exchange stimulated by spontaneous complex formation with cationic comb-type copolymer. J Am Chem Soc 124(43): 12676–12677
84. Wu L et al (2008) Poly(L-lysine)-graft-dextran copolymer accelerates DNA hybridization by two orders. Soft Matter 4:744–747
85. Moriyama R et al (2011) The role of cationic comb-type copolymers in chaperoning DNA annealing. Biomaterials 32(30):7671–7676

86. Darisipudi MN et al (2011) Dual blockade of the homeostatic chemokine CXCL12 and the proinflammatory chemokine CCL2 has additive protective effects on diabetic kidney disease. Am J Pathol 179(1):116–124
87. Urata H et al (1992) Synthesis and properties of mirror-image DNA. Nucleic Acids Res 20(13):3325–3332
88. Miyoshi D, Nakao A, Sugimoto N (2002) Molecular crowding regulates the structural switch of the DNA G-quadruplex. Biochemistry 41(50):15017–15024
89. Xue Y et al (2007) Human telomeric DNA forms parallel-stranded intramolecular G-quadruplex in K + solution under molecular crowding condition. J Am Chem Soc 129(36): 11185–11191
90. Kumar N, Basundra R, Maiti S (2009) Elevated polyamines induce c-MYC overexpression by perturbing quadruplex-WC duplex equilibrium. Nucleic Acids Res 37(10):3321–3331
91. Yin F, Liu J, Peng X (2003) Triethylene tetraamine: a novel telomerase inhibitor. Bioorg Med Chem Lett 13(22):3923–3926
92. Petraccone L, Barone G, Giancola C (2005) Quadruplex-forming oligonucleotides as tools in anticancer therapy and aptamers design: energetic aspects. Curr Med Chem Anticancer Agents 5(5):463–475
93. Bock LC et al (1992) Selection of single-stranded DNA molecules that bind and inhibit human thrombin. Nature 355(6360):564–566
94. Li WX et al (1994) A novel nucleotide-based thrombin inhibitor inhibits clot-bound thrombin and reduces arterial platelet thrombus formation. Blood 83(3):677–682
95. Bates PJ et al (1999) Antiproliferative activity of G-rich oligonucleotides correlates with protein binding. J Biol Chem 274(37):26369–26377
96. Girvan AC et al (2006) AGRO100 inhibits activation of nuclear factor-kappaB (NF-kappaB) by forming a complex with NF-kappaB essential modulator (NEMO) and nucleolin. Mol Cancer Ther 5(7):1790–1799
97. Wyatt JR et al (1994) Combinatorially selected guanosine-quartet structure is a potent inhibitor of human immunodeficiency virus envelope-mediated cell fusion. Proc Natl Acad Sci USA 91(4):1356–1360
98. Kusser W (2000) Chemically modified nucleic acid aptamers for in vitro selections: evolving evolution. J Biotechnol 74(1):27–38
99. Tran PLT et al (2011) A mirror-image tetramolecular DNA quadruplex. Chem Commun (Camb) 47(19):5437–5439
100. Gros J et al (2007) Guanines are a quartet's best friend: impact of base substitutions on the kinetics and stability of tetramolecular quadruplexes. Nucleic Acids Res 35(9):3064–3075
101. Gros J et al (2008) 8-Amino guanine accelerates tetramolecular G-quadruplex formation. Chem Commun (Camb) 25:2926–8
102. Sannohe Y, Sugiyama H (2010) Overview of formation of G-quadruplex structures. Curr Protoc Nucleic Acid Chem 40:Unit 17.2.1–17.2.17, Chapter 17
103. Saccà B, Lacroix L, Mergny JL (2005) The effect of chemical modifications on the thermal stability of different G-quadruplex-forming oligonucleotides. Nucleic Acids Res 33(4): 1182–1192
104. Randazzo A et al (2004) NMR solution structure of a parallel LNA quadruplex. Nucleic Acids Res 32(10):3083–3092
105. Nielsen JT, Arar K, Petersen M (2006) NMR solution structures of LNA (locked nucleic acid) modified quadruplexes. Nucleic Acids Res 34(7):2006–2014
106. Petraccone L et al (2006) Energetic aspects of locked nucleic acids quadruplex association and dissociation. Biopolymers 83(6):584–594
107. Esposito V et al (2009) Effects of the introduction of inversion of polarity sites in the quadruplex forming oligonucleotide TGGGT. Bioorg Med Chem 17(5):1997–2001
108. Datta B, Schmitt C, Armitage BA (2003) Formation of a PNA2-DNA2 hybrid quadruplex. J Am Chem Soc 125(14):4111–4118

109. Paul A et al (2008) Combining G-quadruplex targeting motifs on a single peptide nucleic acid scaffold: a hybrid (3+1) PNA-DNA bimolecular quadruplex. Chemistry 14(28):8682–8689
110. Roy S et al (2011) Kinetic discrimination in recognition of DNA quadruplex targets by guanine-rich heteroquadruplex-forming PNA probes. Chem Commun (Camb) 47(30):8524–8526
111. Krishnan-Ghosh Y, Stephens E, Balasubramanian S (2004) A PNA4 quadruplex. J Am Chem Soc 126(19):5944–5945
112. Esposito V et al (2003) PNA-DNA chimeras forming quadruplex structures. Nucleosides Nucleotides Nucleic Acids 22(5–8):1681–1684
113. Wlotzka B et al (2002) In vivo properties of an anti-GnRH Spiegelmer: an example of an oligonucleotide-based therapeutic substance class. Proc Natl Acad Sci USA 99(13): 8898–8902
114. Kim Y, Yang CJ, Tan W (2007) Superior structure stability and selectivity of hairpin nucleic acid probes with an L-DNA stem. Nucleic Acids Res 35(21):7279–7287
115. Lin C et al (2009) Mirror image DNA nanostructures for chiral supramolecular assemblies. Nano Lett 9(1):433–436
116. Sayyed SG et al (2009) Podocytes produce homeostatic chemokine stromal cell-derived factor-1/CXCL12, which contributes to glomerulosclerosis, podocyte loss and albuminuria in a mouse model of type 2 diabetes. Diabetologia 52(11):2445–2454
117. Garbesi A et al (1993) L-DNAs as potential antimessenger oligonucleotides: a reassessment. Nucleic Acids Res 21(18):4159–4165
118. Urata H et al (2002) Anti-HIV-1 activity of L-DNA quadruplex. Nucleic Acids Res Suppl (2):163–164
119. Szalai VA, Singer MJ, Thorp HH (2002) Site-specific probing of oxidative reactivity and telomerase function using 7,8-dihydro-8-oxoguanine in telomeric DNA. J Am Chem Soc 124(8):1625–1631
120. Esposito V et al (2004) Effects of an 8-bromodeoxyguanosine incorporation on the parallel quadruplex structure [d(TGGGT)]4. Org Biomol Chem 2(3):313–318
121. Virgilio A et al (2005) 8-Methyl-2′-deoxyguanosine incorporation into parallel DNA quadruplex structures. Nucleic Acids Res 33(19):6188–6195
122. Mekmaysy CS et al (2008) Effect of O6-methylguanine on the stability of G-quadruplex DNA. J Am Chem Soc 130(21):6710–6711
123. Matsugami A et al (2007) Structure of a human telomeric DNA sequence stabilized by 8-bromoguanosine substitutions, as determined by NMR in a K+ solution. FEBS J 274(14): 3545–3556
124. Tran PLT et al (2011) Effects of 8-methylguanine on structure, stability and kinetics of formation of tetramolecular quadruplexes. Biochimie 93(3):399–408
125. Amato J et al (2011) Targeting G-quadruplex structure in the human c-Kit promoter with short PNA sequences. Bioconjug Chem 22(4):654–663
126. Stefl R et al (2003) Formation pathways of a guanine-quadruplex DNA revealed by molecular dynamics and thermodynamic analysis of the substates. Biophys J 85(3):1787–1804
127. Marsh TC, Henderson E (1994) G-wires: self-assembly of a telomeric oligonucleotide, d(GGGGTTGGGG), into large superstructures. Biochemistry 33(35):10718–10724
128. Kotlyar AB et al (2005) In vitro synthesis of uniform poly(dG)-poly(dC) by Klenow exo-fragment of polymerase I. Nucleic Acids Res 33(2):525–535
129. Cohen H et al (2007) Polarizability of G4-DNA observed by electrostatic force microscopy measurements. Nano Lett 7(4):981–986
130. Lyonnais S et al (2008) Functionalization of DNA G-wires for patterning and nanofabrication. Nucleic Acids Symp Ser (Oxf) (52):689–690
131. Protozanova E, Macgregor RB Jr (1996) Frayed wires: a thermally stable form of DNA with two distinct structural domains. Biochemistry 35(51):16638–16645
132. Batalia MA et al (2002) Self-assembly of frayed wires and frayed-wire networks: nanoconstruction with multistranded DNA. Nano Lett 2(4):269–274
133. Venczel EA, Sen D (1996) Synapsable DNA. J Mol Biol 257(2):219–224

134. Fahlman RP, Sen D (1998) Cation-regulated self-association of "synapsable" DNA duplexes. J Mol Biol 280(2):237–244
135. Kaucher MS, Harrell WA Jr, Davis JT (2006) A unimolecular G-quadruplex that functions as a synthetic transmembrane Na+ transporter. J Am Chem Soc 128(1):38–39
136. Sakai N et al (2006) Dendritic folate rosettes as ion channels in lipid bilayers. J Am Chem Soc 128(7):2218–2219
137. Ma L et al (2008) Large and stable transmembrane pores from guanosine-bile acid conjugates. J Am Chem Soc 130(10):2938–2939
138. Lubitz I, Kotlyar A (2011) Self-assembled G4-DNA-silver nanoparticle structures. Bioconjug Chem 22(3):482–487
139. Lubitz I, Kotlyar A (2011) G4-DNA-coated gold nanoparticles: synthesis and assembly. Bioconjug Chem 22(10):2043–2047
140. Miyoshi D et al (2005) DNA nanowire sensitive to the surrounding condition. Nucleic Acids Symp Ser (Oxf) (49):43–44
141. Miyoshi D et al (2007) Artificial G-wire switch with 2,2′-bipyridine units responsive to divalent metal ions. J Am Chem Soc 129(18):5919–5925
142. Mergny JL et al (2006) Quadruplex Nucleic Acids. Royal Society of Chemistry Chapter 2. Energetics, Kinetics and Dynamics of Quadruplex Folding:31–80

Index

A
Accessible surface areas, 189
Acriflavine, 165
Aggregation-induced emission (AIE), 137
Alexa Fluor 549, 168
8-Aminoguanine, 263
2-Aminopurine (2-AP), 219
Ammonium-porphyrazine, 153
Anthracyclines, 164
Anthraquinone derivatives, 113, 136
Anticancer, 1, 113, 164, 212
Association, kinetics, 243
  process analysis, 249
Azatrux, 159, 231

B
Bead models, 28, 40, 179, 186
Benzimidazole derivatives, 125
Benzoindoloquinoline derivatives, 144
Berberine, 140
Biotechnology applications, 243
Bis(vinylpyridinium)carbazole, 134
Bis(benzimidazole) derivatives, 125
Bis(pyridinium)porphyrin, 136
Bis(benzimidazole)pyridine (bzimpy) platinum(II) complexes, 147
Bis(pyrazole)pyridine (dPzPy) platinum(II) complexes, 147
Bisquinolinum pyridodicarboxamide motif, 167
BMSG-SH3, 255
BMVC, 104, 130
BOQ_1, 159
Borondipyrromethene (BODIPY), 167
BRACO-19, 39, 255
8-Bromoguanine, 185, 263
BSA, 58
BSU6039, 255

C
CA46 allele-specific transcriptional assay, 10
Carbazoles, 130
Cationic comb-type copolymers, 257
Chaperones, 255
Chemical modifications, 243
Circular dichroism, 12, 67, 211, 213, 222
  induced (ICD), 223
Click chemistry, 168
c-MYC (proto-oncogene), 1, 4, 181, 225
  NMR structure, 5
Concentration, 254
Coralyne, 163
Cosolutes, 98
Cryptolepines, 163
Crystal violet (CV), 128
Cu(TMPyP4), 136
Cyanine dyes, 115

D
Dactyloscopy, 52
Daunomycin, 255
7-Deazaguanine, 263
Diazoniadibenzochryzene, 163
Dipyridophenazine (dppz), 145, 147
Distamycin, 234
DNA, 67
  injected, intracellular localization, 54
  junctions, 93
  nanostructures, 265
  polymorphism, 181
  synapsable, 265

DNA (*cont.*)
targeting, 2
telomeric, 23
triplexes, 93
DODC, 122
Doxorubicin, 164, 169
Drug design, 39
Drug discovery, 179
DTCC, 122
DTPA ligands, 156
Duplex–quadruplex competition, 99
Dyes, 114

E
ELISA, 91
End-stacking, 225
Ethanol, 58
Ethidium bromide (EtBr), 126
Europium(III), 156

F
Ficoll, 58, 60
Fluorescein (FAM), 221
Fluorescence, 111, 211
Fluorescence resonance energy transfer (FRET), 221
melting assay, 221
Fluorescence spectroscopy, 219
Fluorescent probes, 111
Folding pathway, 264
Form-3, 24
Frayed wires, 265
F÷rster distance, 221

G
Gene regulation, 1
Glycerol, 58, 95, 130
Glycosidic bond angle (GBA), 67, 71
GQC-05, 10, 12
GQN1, 255, 266
G-quadruplexes (G4), 1
DNAs, 68
drug interactions, 211
higher-order, 23
physiological conditions, 55
G-quadruplex-forming sequences (GPQSs), 2
G-quartet, 245
Groove binding, 234
Guanine base modifications, 263
Guanosine 5'-monophosphates (5'-GMP), 244
G-wires, 265

H
Heptads, 248
Hexads, 248
Higher-order structure, 23
HIV-1, 259
nucleocapsid protein (NCp), 255
Hoechst 33258, 104, 116, 125, 153
Hydration, 58, 87, 102
Hydrodynamics, 179

I
Induced circular dichroism (ICD), 223
Induced intercalated triad pocket, 6
In vivo, 47
Ion channels, 266
Ionic strength, 253
Isaindigotone, 142
Isoquinoline alkaloids, 140, 163
Isothermal titration calorimetry (ITC), 211, 214
thermodynamic analysis, 215

L
Lanthanides, 156
L1BOD-7OTD, 168
L1H1-7OTD, 168
Ligands, 114, 211
Light-off probes, 158
Light-up probes, 111
Locked nucleic acids (LNA), 259
Loop-interacting, 225

M
Malachite green (MG), 128
Metal complexes, 145
Methylene blue, 165
Methyl green (MEG), 128
8-Methylguanine, 185, 263
Molecular crowding, 47, 58, 87, 88, 95
in vitro, 90
in vivo, 89
reagents, 91
thermodynamics, 100
Molecular dynamics (MD), 28
Molecular modeling, 23
MOQ_2, 159, 160
Multi-quadruplex structures, 26
energetics, 35

N
Nanoparticles, 266
Nanotechnology, 243

Nanowires, 265
Naphthalene diimide, 255
Naphthalimides, 161
Natural products, 140
Netropsin, 235
N-Methyl-mesoporhyrin IX, 136, 155
NMR, in-cell, 47, 50
Nucleic acids, 50, 111, 179, 180

O
Octads, 248
2'-*O*-Methyl ribonucleotide (O-RNA), 259
Oncogenes, 181
8-Oxoguanine, 263
Oxytricha nova, 96, 129

P
Palmatine, 140
Papaverine, 143
Paramagnetic electron double resonance (PELDOR) spectroscopy, 62
Pentads, 248
Peptide nucleic acid (PNA), 261
Perylene diimide derivatives, 164
pH, 254
Phthalocyanines, 153
PIPER, 255
Platinum(II) complexes, 145
PNA, 261
Poisson–Boltzmann surface area calculations (MM–PBSA), 30
Poly(ethylene glycol) (PEG), 58, 60, 91, 257
Polyamines, 257
Polymers, cationic comb-type copolymers, 257
Polymorphism, 95
Porphyrin-anthraquinone dyads, 135
Porphyrins, 134
  cationic, 225
Potassium, 7, 31, 34, 49, 56, 59, 129, 131, 147, 198, 253
Potential quadruplex sequences (PQS), 48
Proflavine, 165
Protoporphyrin IX (PPIX), 136
Pyridinium-porphyrazine, 153
Pyridodicarboxamide (PDC), 117
Pyridyloxazoles, 162
Pyrrolo[2,1-*b*]quinazoline, 142

Q
Quadruplex DNA, 111, 211
  probes, permanent (tagged), 167
Quadruplexes, mirror image, 262
  tetramolecular, 243
Quarfloxin, 14
Quartets, 245
Quercetin, 142
Quinacridine derivatives, 160
Quindoline-i:c-MYC G-quadruplex, 7

R
Restriction of intramolecular rotation (RIR), 137
RHSP4, 255
Ribozyme, 95
RNA, 67, 93
  molecular crowding, 93
  polymorphism, 181
  2'-sugar-modified analogs, 259
  telomeric, 23
RNA polymerase II, DNA-dependent, 39
Ruthenium(II) complexes, 147
Rutin, 143

S
Sabarubicin, 164
Sedimentation, 179
Sequence, 254
Singular value decomposition (SVD), 35
Solvent accessibility surface area (SASA), 33
Strand polarity modifications, 261
Structural polymorphism, 87
Surface plasmon resonance (SPR), 91
SYBR Green I, 120

T
TASPI-TBA, 166
Telomerase inhibition, 212
Telomeres, 2, 181
Telomestatin, 7
  7OTD, 167
Temperature, 251
Terbium(III), 156
Tetraazoniapentaphenopentaphene, 163
Tetrakis(1-methyl-4-pyridyl)porphyrine (TMPyP4), 7, 134, 225
Tetrakis[4-(2-triethylammonioethoxy)phenyl] ethene (TTAPE), 137
Tetramethylrhodamine (TAMRA), 221
Thermal stability, 243, 251
Thermodynamics, 87, 100
  stability, 23
Thiazole Orange (TO), 115, 220
6-Thioguanine, 185

Thrombin-binding aptamer (TBA) quadruplex, 166, 259
TMe_2D_4, 136
TMPyP4, 7, 134, 225, 255
Topoisomerase II, 255
TO-PRO-3, 120
TOxaPy, 159, 162
Transition-metal complexes, 145
Transition-metal phthalocyanines, 153
Transmembrane ion channels, 266
Triazatruxene (azatrux), 39
Triethylene tetraamine, 257
Trimethylamine oxide (TMAO), 94
Triphenylmethane dyes, 128
Tris[4(1-piperidino)butyl]diindolocarbazole (azatrux), 159, 231
TrPyP4, 134, 135
TTHA ligands, 156
Twisted intramolecular charge transfer (TICT), 128

W
Water activity, 87

X
Xenopus laevis, 47, 50, 53, 98

Z
Zn-DIGP, 153