

NMR investigations of G-quadruplex structures

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich Biochemie, Chemie und Pharmazie
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von

Irene Bessi

aus Pistoia, Italien

Frankfurt (2016)

(D 30)

vom Fachbereich Biochemie, Chemie und Pharmazie der
Johann Wolfgang Goethe - Universität als Dissertation angenommen.

Dekan: Prof. Dr. Michael Karas

Gutachter: Prof. Dr. Harald Schwalbe
Prof. Dr. Alexander Heckel
Prof. Dr. Janez Plavec

Datum der Disputation: 29.06.2016

To my parents

This thesis was prepared under the supervision of Prof. Dr. Harald Schwalbe between January 2012 and January 2016 at the Institute for Organic Chemistry and Chemical Biology of the Goethe-University Frankfurt am Main.

Table of contents

Table of contents	i
List of abbreviations.....	iii
Overview and summary	1
1 DNA from A to Z	7
1.1 DNA building blocks.....	8
1.2 Canonical base pairing.....	10
1.3 Non-canonical base pairing	13
2 G-quadruplex structures	17
2.1 Cation coordination to G-quadruplex.....	18
2.2 Stacking of G-tetrads	23
2.3 Polymorphism of G-quadruplex	24
2.4 Polymorphism of human telomeric DNA G-quadruplex.....	28
3 Location of G-quadruplexes in the genome and biological function	33
3.1 G-quadruplex structures at telomeres	35
3.2 G-quadruplex structures in transcription	42
3.3 G-quadruplex structures in replication	48
3.4 G-quadruplexes at other locations.....	49
4 G-quadruplex as novel anti-cancer target.....	51
4.1 Targeting DNA G-quadruplexes with small molecules	52
4.2 Occurrence and regulation by RNA G-quadruplexes.....	56
5 NMR of G-quadruplexes.....	59
5.1 Favouring a single G-quadruplex conformation	59
5.2 NMR parameters: chemical shifts and coupling constants	61
5.3 Assignment strategies	65
5.4 Binding and dynamics of cations studied by NMR	70
5.5 Real-time NMR to monitor G-quadruplex folding.....	71
6 G-quadruplex ligands	75
6.1 Research article: Spectroscopic, molecular modelling, and NMR-spectroscopic investigation of the binding mode of the natural alkaloids berberine and sanguinarine to human telomeric G-quadruplex DNA	85
6.2 Research article: A nucleus-imaging probe that selectively stabilizes a minor conformation of <i>c-MYC</i> G-quadruplex and down-regulates <i>c-MYC</i> transcription in human cancer cells	99

6.3	Research article: A fluorescent guanosine dinucleoside as a selective switch-on sensor for <i>c-myc</i> G-quadruplex DNA with potent anticancer activities	117
7	Insight into G-quadruplex folding kinetics by real-time NMR	125
7.1	Research article: Involvement of long-lived intermediate states in the complex folding pathway of the human telomeric G-quadruplex.....	129
8	G-quadruplex in nanotechnology	137
8.1	Research article: Photoresponsive Formation of an Intermolecular Minimal G-Quadruplex Motif	141
	German summary	149
	Bibliography.....	155

List of abbreviations

°C	temperature in degrees Celsius
1D, 2D, ..., nD	one-, two-dimensional, ..., n-dimensional
2'-dG	2'-deoxyguanosin
2'-F-araG	2'-deoxy-2'-fluoro-arabinoguanosine
2'-F-r	2'-deoxy-2'-fluoro-guanosine
5'-UTR	5'-untranslated region
A	adenine
Az	azobenzene
bis-tris	2,2-bis(hydroxymethyl)-2,2"-nitrilotriethanol
bp	base pair
<i>B. mori</i>	<i>Bombyx mori</i>
BSA	bovine serum albumin
C	cytosine
C ⁺	protonated cytosine
CD	circular dichroism
<i>C. elegans</i>	<i>Caenorhabditis elegans</i>
COSY	corrleated spectroscopy
CSP	chemical shift perturbation
d-	desoxyribo-
δ	chemical shift or torsion angle
D ₂ O	deuterated water
DNA	desoxyribonucleic acid
ds-	double strand
DSS	3-(trimethylsilyl)-1-propanesulfonic acid
<i>e.g.</i>	<i>exempli gratia</i> , for example
EM	electron microscopy
EPR	electron paramagnetic resonance
<i>et al.</i>	<i>et alia</i> , and others
EXSY	NMR exchange spectroscopy
FRET	Förster resonance energy transfer
γ	gyromagnetic ratio (rad·s ⁻¹ ·T ⁻¹) or torsion angle
G	guanine
G4P	genome sequence with potential for G-quadruplex formation (according to Maizels)
h	hour
HMBC	heteronuclear multiple bond correlation spectroscopy
HMQC	heteronuclear multi-quantum correlation spectroscopy
H ₂ O	water
<i>H. sapiens</i>	<i>Homo sapiens</i>
HSQC	heteronuclear single-quantum correlation spectroscopy
<i>i.e.</i>	<i>id est</i> , that is
J	scalar coupling constant
K	temperature in degrees Kelvin
K ⁺	potassium cation
kb	kilobase
KCl	potassium chloride
LNA	locked nucleic acids
M	molar
m	meter
Mg ²⁺	magnesium cation
MHz	Megahertz
min	minute
mRNA	messenger RNA

mtDNA	mitochondrial DNA
Na ⁺	sodium cation
NH ₄ ⁺	ammonium cation
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
nt	nucleotide
N _z	longitudinal nitrogen magnetization
N _z Ex-HSQC	N _z chemical exchange heteronuclear single quantum coherence spectroscopy
P	phase angle of the pseudorotation
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
pH	the negative of the logarithm to base 10 of the hydrogen ion activity (M)
ppm	parts per million
PQS	putative G-quadruplex forming sequences (according to Huppert)
Pu	purine (A or G)
Py	pyrimidine (C or T)
r-	ribo-
R _{1p}	rotating frame relaxation rate
RDC	residual coupling constants
R _g	gyration radius
RNA	ribonucleic acid
rRNA	ribosomal RNA
s	second
sm-FRET	single-molecule Förster resonance energy transfer
S/N	signal-to-noise ratio
SOFAST	band-selective optimized flip-angle short-transient
ss-	single strand
^{tel} IC ₅₀	telomerase half maximal inhibitory concentration
TBA	thrombin binding aptamer
TFO	triplex-forming oligonucleotides
TRAP	telomerase repeat amplification protocol
tRNA	transfer RNA
τ _m	amplitude of the pseudorotation
U	uracil
UNA	unlocked nucleic acids
UV	ultraviolet
Vis	visible
WC	Watson-Crick
%	percent

Overview and summary

This thesis deals with the NMR characterization of the structure and the folding dynamics of DNA G-quadruplexes as potential therapeutic target in cancer therapy and building block for DNA-based nanotechnology.

In presence of stabilizing cations such as Na^+ or K^+ , G-rich sequences can form G-quadruplex structures by stacking of multiple G-tetrads on top of each other, each G-tetrad resulting from the cyclic arrangement of four guanine bases stabilized by Hoogsteen hydrogen bonds. The canonical B-DNA has been regarded for a long time as the only physiologically relevant DNA conformation; however, the role of G-quadruplex structures as regulatory elements in many biological processes has recently been recognized.

The telomeric DNA is formed by the repetition of a G-rich motif and its single-stranded 3'-overhang is the substrate of the telomerase, which is up-regulated in 85-90% of all human cancers. Evidence for the formation of G-quadruplexes at the 3'-end of the human telomeric DNA has been reported. The formation of G-quadruplex structures may affect the recruitment of the telomerase enzyme and the telomere-associated proteins required for the capping and maintenance of the chromosomes end. Therefore many efforts have recently been devoted to investigate small molecules able to bind and stabilize the G-quadruplex structures formed at the telomeres as potential anticancer drugs. However, due to the high polymorphism of the human telomeric DNA, the structural information reported till now on G-quadruplex-ligand complexes is very scarce. To better understand and interpret the binding modes between the G-quadruplex structure and the interacting ligands, more information on the structure and dynamics of the human telomeric DNA is required. A deeper insight into the folding mechanisms of telomeric G-quadruplex is also necessary to elucidate the interaction of the telomeres 3'-overhang with ligands and telomere-binding proteins.

Bioinformatics studies predicted the formation of G-quadruplex structures also in other regions of the genome, suggesting a possible role for G-quadruplex structures in transcription regulation. In particular, the promoter regions of several oncogenes are enriched in potential G-quadruplex forming sequences, compared to tumor suppressor genes and housekeeping genes. The effects of G-quadruplex formation and its stabilization with small molecules have been best characterized for the transcription of the *c-MYC* oncogene, whose expression is deregulated in a number of malignancies. It has been shown that the binding of small molecules to the G-quadruplex structures formed in the G-rich sequence located upstream of the *c-MYC* promoter region results in down-regulation of the *c-MYC* expression. The development of small ligands able to repress *c-MYC* expression is considered an attractive strategy in anti-cancer therapy.

Besides playing an important role in genome maintenance and gene regulation of biological processes, the G-quadruplex structures have been used in nanotechnology to drive self-assembly of nanoarchitectures and build nanomachines.

The first part of this thesis (Chapters 1-5) introduces the reader to the world of G-quadruplexes.

The main features of the classic Watson-Crick double helix and alternative non-B DNA structures are illustrated in **Chapter 1**. Many different base pairing schemes are possible, besides the canonical Watson-Crick motif, thereby expanding the structural complexity of DNA. Non-canonical base pairing, such as Hoogsteen hydrogen bonding, enables the assembly of triplets and quartets, which are the building blocks of triplex and quadruplex structures, respectively.

The structural characteristics of DNA G-quadruplexes are delineated in detail in **Chapter 2**. G-quadruplex structures are extremely polymorphic, in terms of strands orientation, loops geometry, grooves width and arrangement of the glycosidic torsion angles. The various structural elements as well as the different cation coordination geometries are here presented, with a special emphasis on the diversity of conformations reported for the telomeric DNA G-quadruplexes.

Chapter 3 describes the biological roles of G-quadruplex structures in the genome. After introducing the architecture of the telomeric DNA and its interacting proteins, the mechanism of the telomeres elongation catalysed by the telomerase enzyme and its implications for cancer are discussed. The occurrence of G-quadruplex structures in functional regions of the genome, such as promoter regions of oncogenes, and their possible roles in regulating the gene transcription are then outlined in the second part of the chapter.

The potential of G-quadruplex as a novel anti-cancer target is examined in **Chapter 4** and the proposed anti-cancer mechanisms for ligands stabilizing G-quadruplex structures are discussed. RNA G-quadruplexes and their putative role in gene regulation at the level of translation are briefly illustrated at the end of the chapter.

A general overview on the NMR methods to investigate the G-quadruplex structures is presented in **Chapter 5**. The experimental set-up used for the real-time NMR studies of the G-quadruplex folding is also described.

The second part of the thesis (Chapters 6-8), which is the cumulative part, comprises the original publications grouped in three Chapters according to the topic:

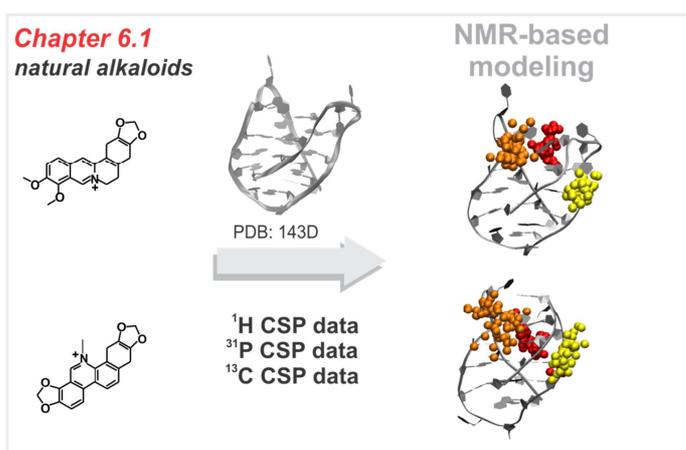
- G-quadruplex interacting ligands (Chapter 6),
- G-quadruplex folding kinetics (Chapter 7),
- G-quadruplex in nanotechnology (Chapter 8).

A brief introduction to the topic opens each chapter, providing the framework for the following papers. The personal contribution of this thesis' author to each publication is stated in the introduction to the respective article.

The state of the art on small molecules targeting G-quadruplex structures is given at the beginning of **Chapter 6**, including a summary of the experimental structures of G-quadruplexes in complex with ligands available up to date.

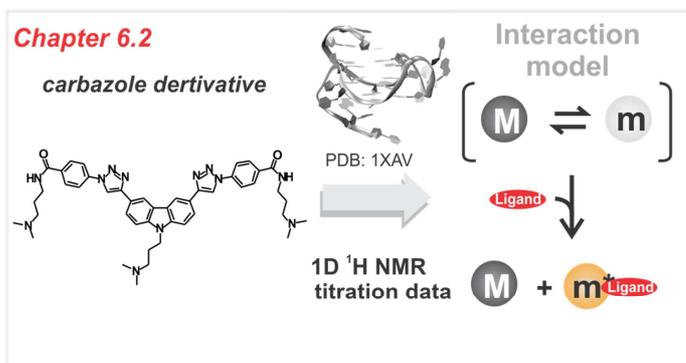
The publications presented in Chapters 6.1-6.3 are concerned with the elucidation of the interaction modes between DNA G-quadruplexes and selected ligands with potential therapeutic applications.

The binding ability of two natural alkaloids (berberine and sanguinarine) to telomeric G-quadruplexes is examined in **Chapter 6.1**. The ability of carbazole and diguanosine derivatives (synthesized in the group of Prof. Dash, IISER, Kolkata) to interact with *c-MYC* G-quadruplex and to down-regulate *c-MYC* expression is explored in **Chapter 6.2** and **Chapter 6.3**, respectively.



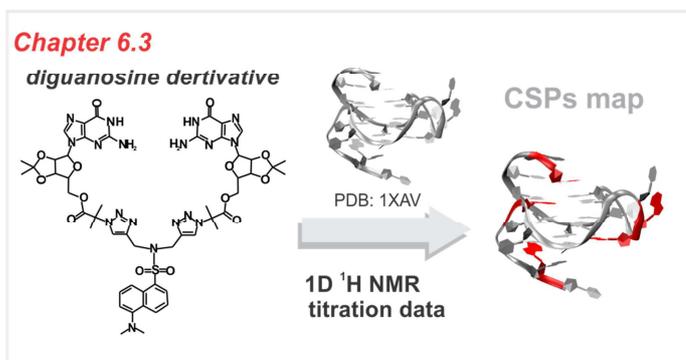
Achievements:

- Identification of a high stoichiometry of binding with ligand self-association induced by the DNA binding
- Mapping of the binding sites via NMR-driven docking



Achievements:

- Proposal of an interaction model (conformational selection) on the basis of 1D ¹H NMR titration data



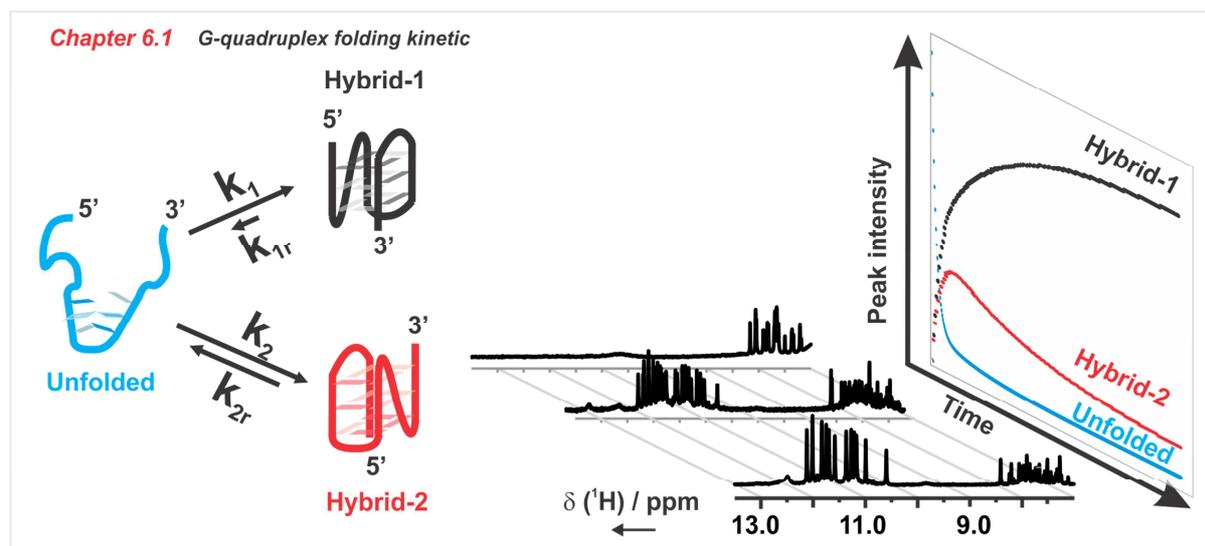
Achievements:

- Mapping of the interaction interface between the ligand and the *c-MYC* G-quadruplex via ¹H chemical shift perturbation (CSP) data.

The energy landscape of human telomeric G-quadruplex structures is discussed in **Chapter 7**, in light of the experimental kinetic studies as well as molecular dynamics simulations reported in literature until now. Up to date there is no general consensus regarding the folding pathway of unimolecular human telomeric G-quadruplex, in particular due to the lack of atomic resolution data on the species involved in the folding. **Chapter 7.1** presents the first real-time NMR study of the human telomeric G-quadruplex folding kinetics.

Achievements:

- Application of real-time NMR to provide atomic resolution insight into the folding pathway of the human telomeric G-quadruplex
- Proposal of a folding mechanism (kinetic partitioning) involving long-lived intermediate states
- Characterization of the folding topology of the species involved in the folding

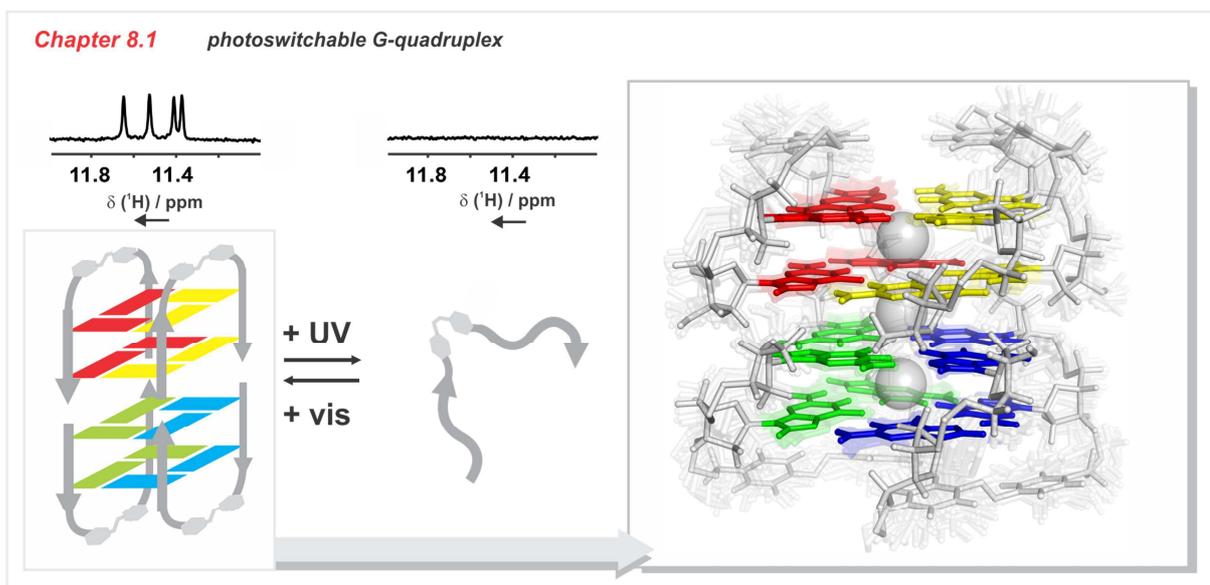


The final chapter of this thesis (**Chapter 8**) outlines the potential of G-quadruplex structures as building blocks in nanotechnology. After illustrating briefly the additional possibilities offered by alternative non-B DNA structures to programme nanomaterials, a number of applications employing G-quadruplex structures in different fields of nanotechnology are described.

The article presented in **Chapter 8.1** investigates the structural and photoswitching properties of a novel intermolecular azobenzene-containing G-quadruplex synthesized in the group of Prof. Heckel (Goethe University, Frankfurt).

Achievements:

- Identification of an azobenzene derivative able to form an homogeneous intermolecular G-quadruplex structure
- Characterization of its photoswitching behaviour
- Determination of the NMR structure of the G-quadruplex.



List of publications presented in the cumulative part

- “Spectroscopic, molecular modeling, and NMR-spectroscopic investigation of the binding mode of the natural alkaloids berberine and sanguinarine to human telomeric G-quadruplex DNA.” Irene Bessi, Carla Bazzicalupi, Christian Richter, Hendrick R. A. Jonker, Krishna Saxena, Claudia Sissi, Matteo Chioccioli, Sara Bianco, Anna Rita Bilia, Harald Schwalbe, Paola Gratteri. ACS Chem Biol., Volume 7, Issue 6, pages 1109-19 (2012).
- “A Nucleus-Imaging Probe That Selectively Stabilizes a Minor Conformation of c-MYC G-quadruplex and Down-regulates c-MYC Transcription in Human Cancer Cells.” Deepanjan Panda, Manish Debnath, Samir Mandal, Irene Bessi, Harald Schwalbe, Jyotirmayee Dash. Sci. Rep. Volume 19, Issue 5, page 13183 (2015).
- “A Fluorescent guanosine dinucleoside as a selective switch-on sensor for c-myc G-quadruplex DNA with potent anticancer activities.” Y. Pavan Kumar, Sudipta Bhowmik, Rabindra N. Das, Irene Bessi, Sushovan Paladhi, Rita Ghosh, Harald Schwalbe, Jyotirmayee Dash. Chem. Eur. J., Volume 19, Issue 35, pages 11502–11506 (2013).
- “Involvement of long-lived intermediate states in the complex folding pathway of the human telomeric G-quadruplex.” Irene Bessi, Hendrick R. A. Jonker, Christian Richter, Harald Schwalbe. Angew. Chem. Int. Ed. Engl., Volume 2, Issue 54(29), pages 8444-8 (2015).
- “Photoresponsive formation of an intermolecular minimal G-quadruplex motif.” Julie Thevarpadam,[#] Irene Bessi,[#] Oliver Binas,[#] Diana P. N. Gonçalves, Chavdar Slavov, Hendrick R. A. Jonker, Christian Richter, Joseph Wachtveitl, Harald Schwalbe, Alexander Heckel.[#] Equal contribution. Angew. Chem. Int. Ed. Engl., in press.

The personal contribution of the author of this thesis to each paper is stated in the cumulative part.

1 DNA from A to Z

A survey of the literature revealed that all but four letters of the English alphabet have already been used to indicate the various structural geometries adopted by DNA.^[1]

Since the proposal of the Watson and Crick model in 1953, the “canonical” B-DNA has been considered the only physiologically relevant conformation for a long time. The B-DNA structural model was based on low-resolution fiber diffraction data and the first high-resolution data validating the Watson-Crick model were available only in 1980 by Drew and Dickerson.^[2] The very first X-ray high-resolution single-crystal structure of a duplex DNA was reported one year before and, surprisingly, was featured by a left-handed Z-DNA conformation.^[3] Since 1979, many of alternative DNA structures were discovered within a decade.^[4]

DNA is indeed more versatile and malleable than the Watson-Crick double helix and its structural diversity is determined not only by minor differences in the helical parameters, but also by major structural characteristics like the number of strands or the handedness and the base pairing scheme.

These alternative, non-B DNA structures may play major roles in genome maintenance (*e.g.*, DNA replication, transcription, recombination and repair) and may be involved in cancer and diseases.^[5]

The capability of DNA to adopt various conformations has been also exploited in nanotechnology to build nanomachines and to guide self-assembly.^[6]

1.1 DNA building blocks

Nucleic acids are polymers of nucleotides, each composed of a base, a pentose sugar (ribose in RNA or 2'-deoxyribose in DNA) and a phosphodiester group. Two sequential nucleotides are condensed via a phosphodiester linkage that connects the O3' of a nucleotide to the O5' of the following one.

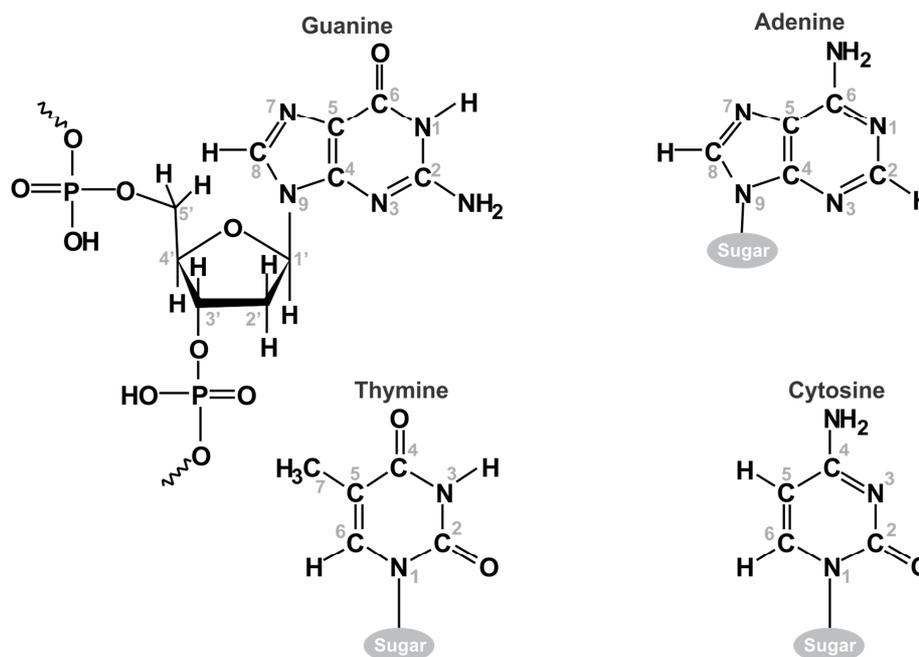


Figure 1 DNA building blocks with atomic numbering and name of each base. The complete nucleotidic unit is shown only for guanine as example, while a schematic structure of the nucleoside is displayed for the other bases.

The pentose sugar is non-planar (puckering) and can adopt various conformations that can be described using the pseudorotation parameters P and τ_m defined in **Figure 2a** as function of the five endocyclic torsion angles (ν_i , $i=1, \dots, 5$), according to the formalism by Altona and Sundaralingam.^[7] The continuum of sugar pucker can be represented on the so-called pseudorotation wheel (**Figure 2b**) using amplitude (τ_m) and phase angle (P) of the pseudorotation. The preferred geometries observed in crystal structures of isolated nucleosides and nucleotides are north conformers C3'-*endo* type (dark gray region) and south conformers C2'-*endo* type (light gray region). The phosphodiester backbone can be described with six torsion angles indicated in **Figure 2c** with α , ..., ζ . These torsion angles can assume restricted values and are inter-correlated, pointing also at a correlated backbone flexibility.^[8] Due to structural constraints, the glycosidic torsion angle χ has a marked preference for two conformational domains: *syn* and *anti* (**Figure 2d**). There is also a correlation between the glycosidic torsion angle and the sugar pucker: a C3'-*endo* sugar is usually characterized by an *anti* glycosidic torsion angle, while residues with a *syn* glycosidic torsion angle prefer to adopt a C2'-*endo* pucker.

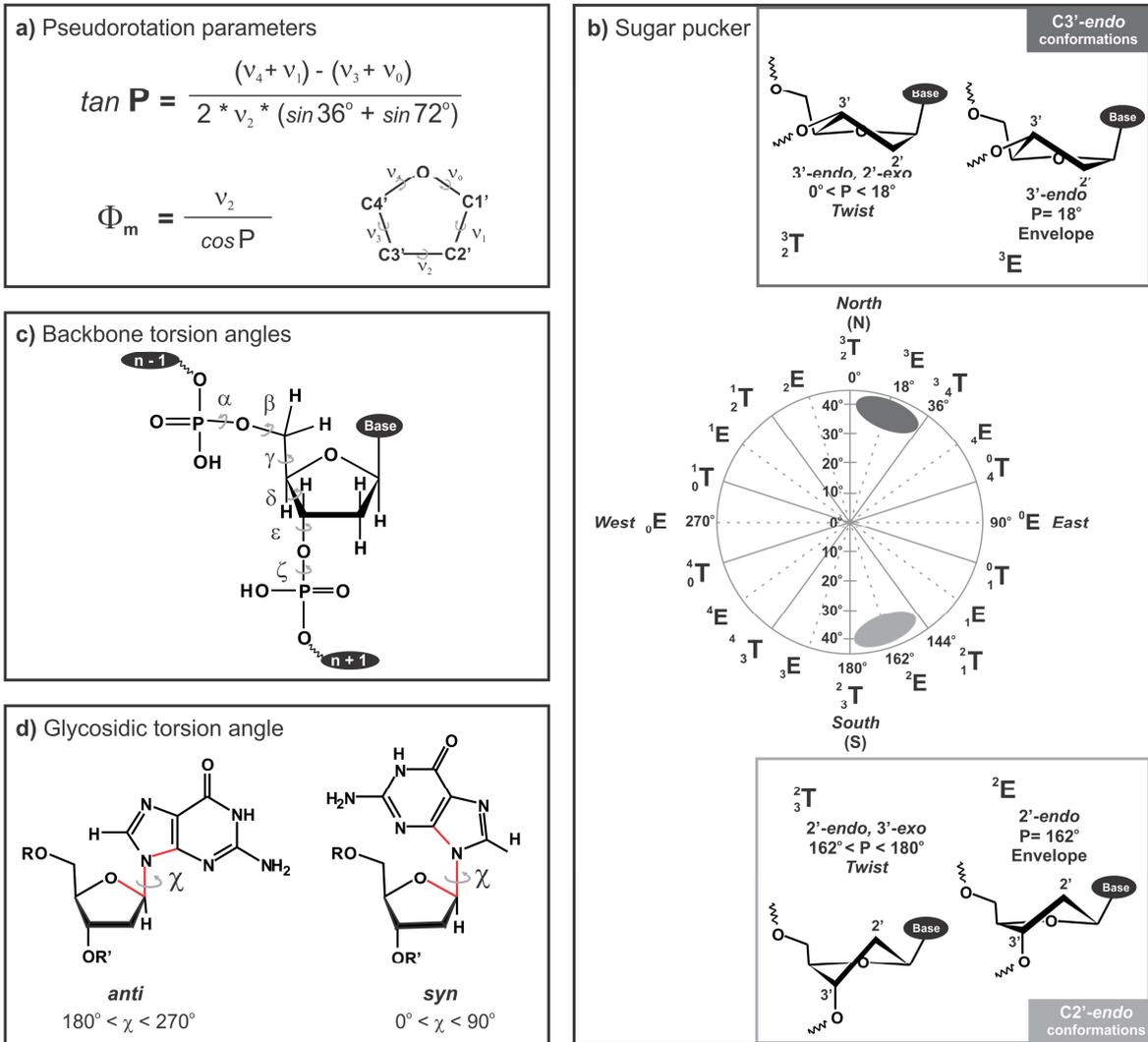


Figure 2 Geometric descriptors for nucleotide torsion angles. a) Definition of pseudorotation parameters and b) pseudorotation wheel for deoxyribose. Preferred conformations of the sugar pucker are south (S), around P=162° (C2'-endo, light gray), or north (N), around P=18° (C3'-endo, dark gray). c) Backbone torsion angles indicated with arrows on a nucleotide unit. d) *anti* and *syn* domains of glycosidic torsion angle χ are defined in terms of the atoms O4'-C1'-N9-C4 in purines (red) and O4'-C1'-N1-C2 in pyrimidines.

1.2 Canonical base pairing

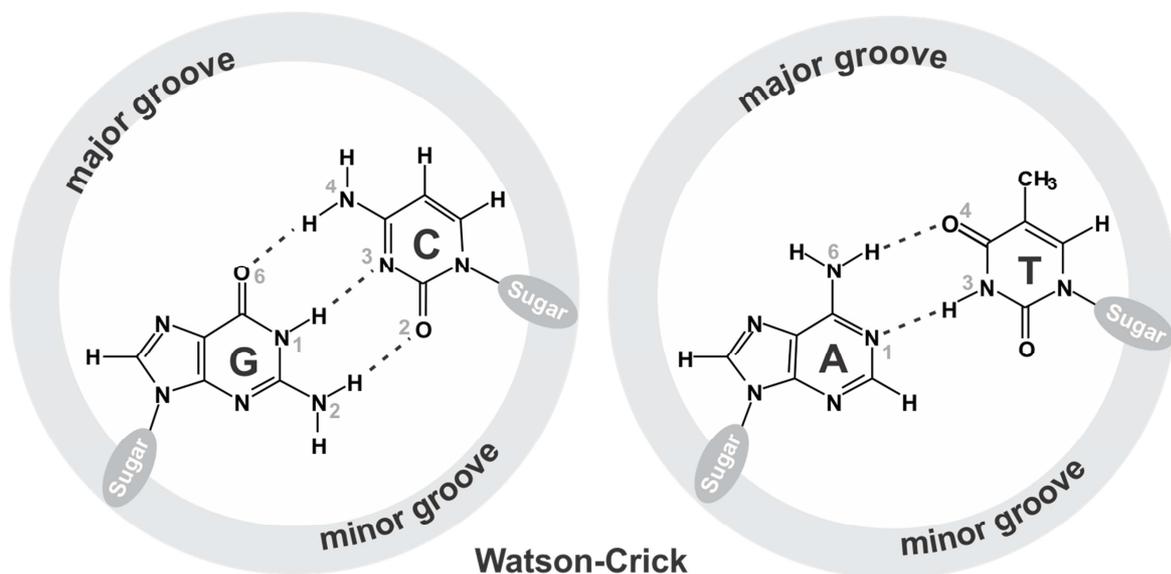


Figure 3 Watson-Crick base pairing pattern and definition of minor groove (the side of the base pair facing the phosphate backbone) and major groove (the other side). The major groove contains O6 and N6 substituents in purines and O4, N4 substituents in pyrimidines.

The right-handed B-DNA double helix originally proposed by Watson and Crick is the most common conformation adopted by DNA in the genome under physiological conditions. Watson and Crick postulated a double helical structure with two strands twisted around a common axis anti-parallel to each other in a right-handed fashion (helical twist $\sim 36^\circ$). In this model the strands are held together by a specific hydrogen bonding pattern between purine bases (Pu) and pyrimidine bases (Py), namely adenine pairs with thymine and guanine with cytosine, which immediately suggested a possible mechanism for copying the genetic material.^[9] This is commonly indicated as Watson-Crick (WC) or canonical base pairing scheme (**Figure 3**). The B-DNA structure features ~ 10 base pairs per helical turn and a helical rise of $\sim 3.4 \text{ \AA}$. The sugar pucker of the deoxyribose is mostly in C2'-endo conformation and the geometry of the glycosidic torsion angle is *anti*.^[2]

Depending on salt concentration, hydration level and sequence composition, the classic right-handed B-DNA can maintain the canonical base pairing pattern and, nevertheless, adopt different structures.

The A-DNA form is the best characterized example of a right-handed double helix deviating from B-DNA. The A-DNA is formed in DNA fibers under low humidity conditions and in solution at low water activity, in presence of alcohols. A-DNA is a right handed double helix with helical twist of $\sim 31^\circ$ and ~ 11 base pairs per helical turn. The glycosidic torsion angles are all *anti* and the conformation of the sugar pucker is C3'-endo.

While in B-DNA the base pairs are basically stacked on top of each other in a parallel fashion with small slide, A-DNA features a large slide and the base pairs are tilted and rolled. The helical rise is

consequently reduced to ~ 2.5 Å. The high x-displacement value characteristic of the A-DNA is related to the “hole” in the top view and makes the diameter of A-DNA overall larger compared to the B-DNA (**Figure 4**). Consequently, the minor groove is wide and shallow and the major groove is narrow and deep. The A-conformation is also the one adopted by RNA duplexes. Interestingly, A-DNA is formed in the nascent template-primer duplex in the DNA polymerase active site as result of distortion and underwinding of B-DNA.^[10] The shallow, wide minor groove in A-DNA allows the polymerase for a better reading of the minor groove structure, which is more sensitive to mismatching than the major groove. Therefore, A-DNA formation in the polymerase active site provides an extra mechanism to guarantee correct base-pairing and high fidelity of replication.^[5] Other right-handed variants of the B-DNA form have been reported (B'-DNA, C-DNA, D-DNA).

The first high resolution X-ray crystal structure of a duplex DNA was obtained in 1979 by Wang *et al.* and provided the first striking example of non right-handed DNA structure.^[3] A regular alternation of purines (Pu) and pyrimidines (Py) can adopt a zig-zag conformation called Z-DNA. Two different steps can be identified in such a structure: the 5'-Pu-Py-3' step and the 5'-Py-Pu-3' step, with a relative rotation of two adjacent DNA bases (helical twist) of $\sim -9^\circ$ and $\sim -51^\circ$, respectively. This results in a zig-zag phosphodiester backbone, with purines adopting a *syn* conformation and pyrimidines an *anti* conformation, defining a deep and narrow minor groove and no major groove. The repeat unit is therefore a dinucleotide. The Z-DNA has ~ 12 base pairs per helical repeat and is slimmer than a B-DNA.

In physiological cellular conditions the zig-zag backbone produces high electrostatic repulsion between the phosphate groups, which are pushed into the B-DNA conformation. High salt concentrations or polyvalent cations able to decrease the electrostatic repulsion between the phosphodiester groups stabilize the Z-conformation. For this reason, the biological function of Z-DNA has been object of debate.^[11] It has been shown that potential Z-DNA forming regions are non-randomly distributed in the genome and are overrepresented near the transcription initiation site.^[12] The formation of Z-DNA has been also observed under negative supercoiling conditions.^[13] Furthermore, it has been suggested that Z-DNA can be formed transiently as a result of biological activities, such as torsional strain induced by movement of RNA polymerase during transcription. The discovery of Z-DNA binding domains supports as well a biological role of Z-DNA.^[11]

Figure 4 shows examples of A-, B- and Z-DNA structures selected from the Protein Data Bank (PDB), including an analysis of the base steps parameters.

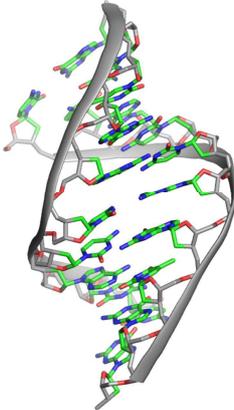
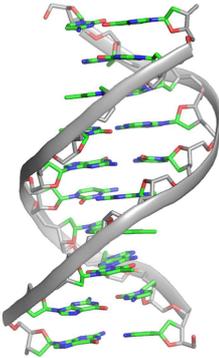
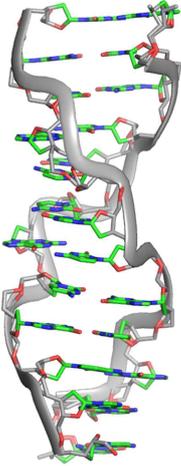
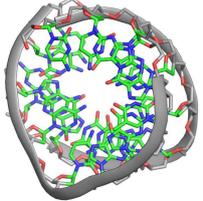
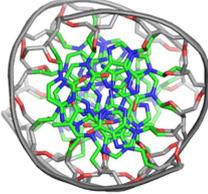
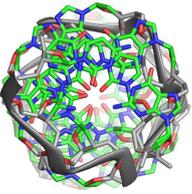
		A-DNA PDB: 1ZF8	B-DNA PDB: 5DNB	Z-DNA PDB: 4OCB	
	Front view				
	Top view				
Base pair step parameters	Shift (Å)	-0.25 (0.84)	0 (0.37)	0 (0.01)	0 (0.27)
	Slide(Å)	-1.76 (0.36)	0.78 (1.22)	5.4 (0.23)	-1.06 (0.22)
	Rise (Å)	3.28 (0.28)	3.31 (0.19)	3.67 (0.32)	3.26 (0.10)
	Tilt (deg.)	-0.13 (3.81)	0 (2.65)	0 (0.84)	0 (0.64)
	Roll (deg.)	5.76 (5.8)	2.22 (6.09)	-2.5 (1.14)	-4.11 (3.39)
	Twist (deg.)	29.41 (4.6)	35.27 (9.27)	-8.26 (0.68)	-50.61 (1.42)
Base pair helical parameters	X-displacement (Å)	-4.49 (1.25)	0.45 (1.92)	-27.02 (5.00)	1.54 (0.50)
	Y-displacement (Å)	0.69 (1.79)	0 (1.04)	0 (3.51)	0 (0.34)
	Helical rise (Å)	2.86 (0.46)	3.22 (0.22)	5.02 (0.72)	3.14 (0.11)
	Inclination (deg.)	11.4 (11.18)	5.1 (9.89)	16.87 (7.24)	4.86 (3.99)
	Tip (deg.)	0.12 (7.39)	0 (4.97)	0 (6.08)	0 (0.76)
	Helical twist (deg.)	30.66 (4.49)	35.9 (9.12)	-8.75 (0.67)	-50.88 (1.30)

Figure 4 Examples of crystal structures (resolution < 2.0 Å) of A-DNA (PDB: 1ZF8), B-DNA (PDB: 5DNB) and Z-DNA (PDB: 4OCB). Orientation and displacement of successive base pairs are reported for each structure according to the guidelines introduced by Dickerson *et al.*^[14] and in a local base-centered reference frame.^[15] Mean values and standard deviations (in brackets) of base pair step parameters and base pair helical parameters result from the analysis of the A-, B- and Z-DNA crystal structures with the software 3DNA.^[16,17] The mean values are in agreement with those reported for a survey of high-resolution crystal structure by Olson *et al.*^[18]

1.3 Non-canonical base pairing

Besides the canonical Watson-Crick base pairs (**Figure 3**), many more different arrangements and base pairing motifs are possible, thereby expanding the structural complexity of DNA. While the reverse Watson-Crick hydrogen bonding pattern (**Figure 5a**) uses the WC edge of the purine base, the Hoogsteen type (**Figure 5b**) employs the so-called Hoogsteen edge, defined by N6 and N7 in adenine and O6 and N7 in guanine.^[19] Homo-purine and homo-pyrimidine base pairs can be also formed (**Figure 5c-d**). These non-canonical base pairs provide additional interaction pattern for building triplets and quartets (**Figure 5e-h**), which are the fundamental structural units of triplex and quadruplex, respectively.

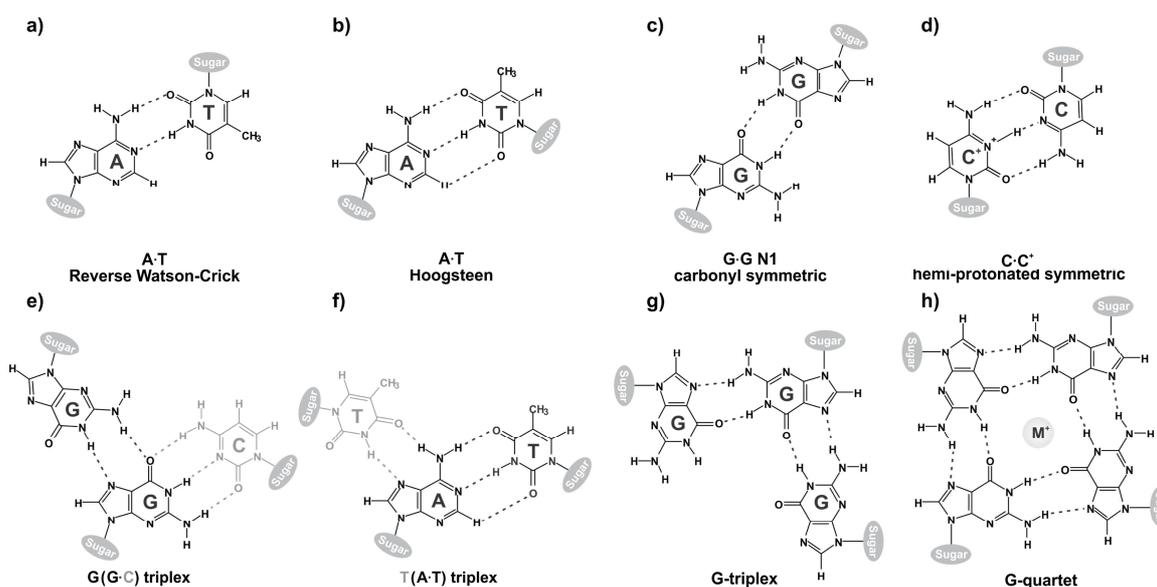


Figure 5 a-d) Selected examples of unusual base pairing schemes. e-g) Base triplets and h) quartet adopting unusual base pairing motifs. In the triplets, the base involved only in canonical WC base pairing is colored in gray.

Non-canonical base pairs can be also formed as transient excited states within a WC double helix with canonical base pairing. In fact, the group of Al-Hashimi has recently detected short-lived, low-populated Hoogsteen base pairs resulting from the flipping of a base in the WC base pair.^[20] Hoogsteen base pairs embedded in a DNA duplex have been observed in complex with proteins but not in naked DNA. The proposed Hoogsteen-WC base pair equilibrium suggests that a transient Hoogsteen base pair could be trapped by protein binding and modulate the binding affinity with a possible functional role.^[21]

Formation of non-canonical base pairing allows expanding the DNA alphabet by the assembly of more than two strands into new structures with unique features very different from the iconic double helical structures.

DNA segments containing mirror repeats in polyPu/polyPy duplexes are called H-palindromes and occur much more frequently in eukaryotic organisms compared to prokaryotes.^[22] A DNA fragment containing one half of the repeat folds back in the major groove of the duplex containing

the mirror repeat and forms an intra-molecular triple helix. The overall structure is called H-DNA and features a triplex and a single-stranded region. H-DNA is formed preferably in negatively supercoiled DNA, since the major groove must unwind in order to be broad enough to accommodate the third strand. Synthetic oligonucleotides have been shown to adopt such a triplex conformation. NMR structures of DNA triplexes are available since 1994 (see DNA-H examples in **Figure 6**), while only one crystal structure containing a pure triplex region flanked by duplexes has been reported until now.^[23]

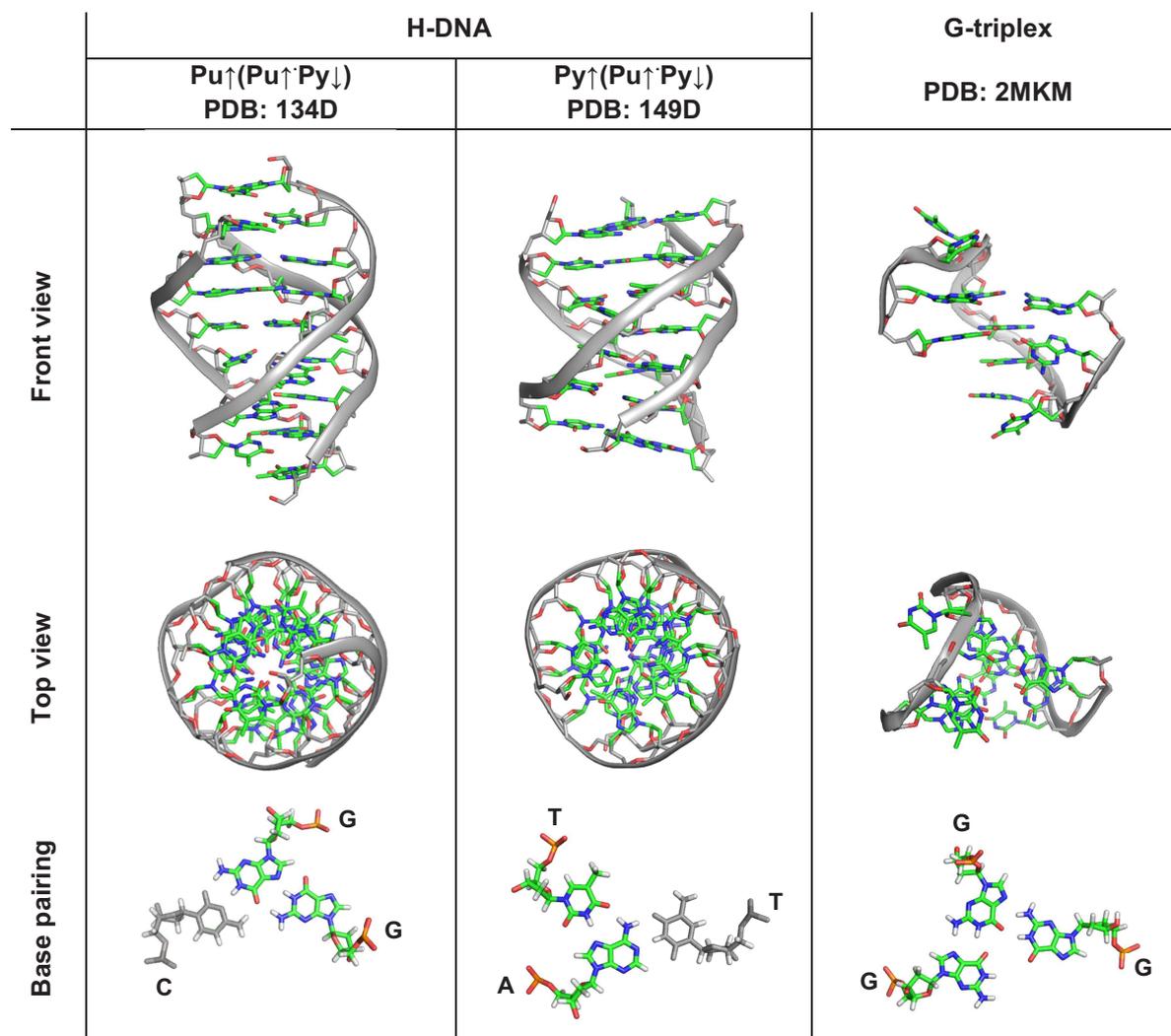


Figure 6 Examples of NMR structure of triplex DNA: Pu(PuPy) (PDB: 134D), Py(PuPy) (PDB:149D) and G-triplex (PDB: 2MKM). Relative polarity of the strands in H-DNA is indicated with arrows, while the Py base of the WC duplex is indicated in gray in the bottom of the scheme.

Two sub-classes of H-DNA can be identified, according to the nature of the third strand: Pu(PuPy) containing triplets such as G(G·C), if the invading strand is purine rich, or Py(PuPy) containing triplets such as T(A·T) and C⁺(G·C), if the invading strand is pyrimidine rich. The third strand always binds to the Hoogsteen edge of the purine in the WC duplex. The formation of a triplex containing C⁺(G·C) triplet is pH-dependent, since the cytosine of the invading strand needs to be protonated. The capability of short, synthetic oligonucleotides to bind the major groove of

duplex DNA to form stable triplexes is the basic principle of the so-called “triplex technology”. Triplex forming oligonucleotides (TFOs) have been used extensively in antigene strategies to inhibit gene transcription and to study mechanisms of DNA repair, damage and mutagenesis.^[24,25]

A G-triplex DNA derived from a truncated sequence of the thrombin binding aptamer (TBA) has recently been characterized by NMR (**Figure 6**, PDB: 2MKM) and it is a plausible intermediate in TBA folding.^[26]

Four-stranded DNA structures such as G-quadruplex and i-motif rely on the formation of non-canonical base pairing schemes in G-rich and C-rich sequences, respectively. The building block of a G-quadruplex is the so-called G-quartet (**Figure 5h**), composed of four guanines arranged cyclically *via* Hoogsteen hydrogen bonds.^[27] In presence of monovalent cations such as Na⁺ or K⁺, more G-tetrads can stack on top of each other and form a G-quadruplex (**Figure 7** and Chapter 2). G-quadruplex forming sequences are widespread throughout the human genome, particularly at the telomeres and at promoter region of oncogenes. The role of G-quadruplex structures in genome maintenance and gene regulation is discussed in more detail in Chapters 3 and 4.

C-rich sequences are able to fold into the i-motif (**Figure 7**) structure.^[28] The basic unit of an i-motif is the symmetric hemi-protonated C⁺·C base pair (**Figure 5d**), which can be formed at acidic pH since it requires the protonation of one cytosine N3. The shared proton jumps between the base-paired cytosines.^[29] An i-motif consists of four strands interacting pairwise *via* intercalated C⁺·C base pairs. Thus, the aromatic moieties of consecutive base pairs are not stacking but rather intercalating. The two interacting strands are parallel between each other and antiparallel to the other two strands, independently on the molecularity of the structure. Such a strands arrangement results in the formation of two wide and shallow major grooves alternated with two narrow minor grooves. Intra-molecular i-motif structures are classified in 5'-end intercalation topology (5'E) or 3'-end intercalation topology (3'E) according to the identity of the outermost C⁺·C base pair.^[30] C-rich sequences able to fold into i-motif structures have been identified in the human telomeres^[31] and in the promoter regions of the oncogenes such as *VEGF*,^[32] *BCL-2*^[33] and *c-MYC*.^[34] All the structures proposed for the i-motifs derived from these oncogene present the structural core described before, but differ in the number of intercalated base pairs (4-7) and in the length of the loops (2-7 nt) connecting the C-stretches. Like G-quadruplexes, i-motifs are supposed to be involved in gene regulation.^[35] Nevertheless, the biological role of i-motif structures has been a matter of some debate because of the acidic pH required for its stability. However, in the case of the *c-MYC* oncogene, formation of the i-motif is supposed to be driven by superhelical stress instead of acidic pH.^[34] Formation of the i-motif as well as G-quadruplexes *in vivo* requires the local unwinding of the B-DNA duplex. It has been proposed that the negative supercoiling induced upstream of the transcription site by the movement of the RNA polymerase (RNAP) is the main force driving the local melting of the duplex and favours the formation of alternative DNA structures.^[34]

The pH-dependent formation of the i-motif is a unique feature that has been widely utilized in nanotechnology to trigger molecular motors, control self-assembly of DNA nanoarchitectures and molecular logic gates.^[36]

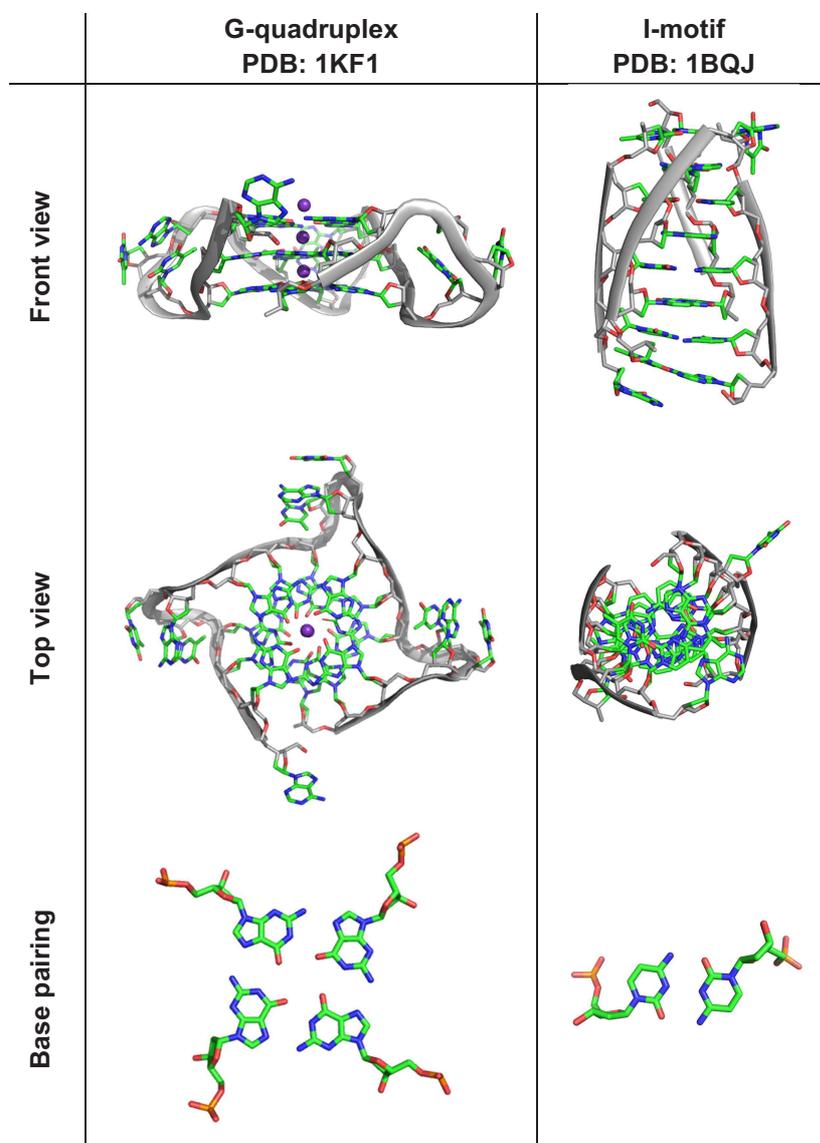


Figure 7 One representative conformation of a G-quadruplex (PDB: 1KF1) and an i-motif (PDB: 1BQJ) crystal structures with G-tetrad (**Figure 5h**) and symmetric hemiprotonated C⁺C base pair (**Figure 5d**) displayed at the bottom. K⁺ stabilizing cations in the G-quadruplex structure are represented with violet spheres.

2 G-quadruplex structures

The peculiar behaviour of guanylic acid (5'-guanosine monophosphate, GMP) to form gels in water at high millimolar concentrations was already reported by the Norwegian chemist Ivar Bang at the beginning of the 20th century.^[37] The structural nature of these hydrogels was explained 42 years later, in 1962, by Gellert, Davies and Lipsett on the basis of X-ray diffraction data on guanylic acid fibers.^[27] They proposed that four 5'-GMP units could interact cyclically in a planar fashion with each guanine using the WC edge to bind the Hoogsteen edge of its neighbour, thereby forming the so-called G-quartet (Figure 8, left). The stacking of more tetrameric units on top of each other (G-quadruplex, Figure 8, right) produces helical structures able to justify the X-ray pattern and gel-like properties of highly concentrated guanylic acid solutions.

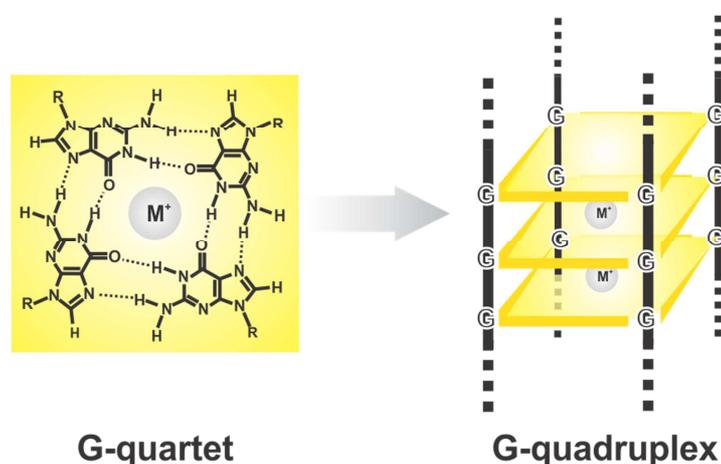


Figure 8 G-quartet (or G-tetrad) unit stabilized by monovalent cations (left) and G-quadruplex structure resulting by the cation-directed stacking of more G-tetrad units.

The essential role of cations in stabilizing G-quartet and template the self-assembly of G-quadruplex structures was reported for the first time by Pinnavaia and co-workers in 1978.^[38] A variety of different hydrophilic and lipophilic guanosine derivatives has been prepared as scaffold to control the self-assembly of materials in water and in organic solvents.^[39]

G-quadruplex structures were initially believed to form only *in vitro* until different groups proposed in 1987 that G-rich sequences found at telomeres and in other functional regions could in fact adopt four-stranded structures with G-tetrads.^[40–42] Since then, a number of potential G-quadruplex forming sequences has been identified in bioinformatics analyses and the structure of these G-quadruplexes, as well as the biological function, have been investigated.

2.1 Cation coordination to G-quadruplex

Cations are necessary for G-quadruplex formation.^[43] As noted in 1980 by Laszlo and co-workers, cation binding and not hydrophobic interactions is the driving force guiding the self-assembly of 5'-GMP: the presence and the nature of the cation is critical to stabilize the stacking of G-tetrads.^[44,45] The guanine carbonyl oxygen atoms (O6) define a planar square in each G-quartet and a bipyramidal antiprism in two G-quartets stacked on top of each other. The stacking of multiple G-quartets forms a cavity which is surrounded by negative charge, given that the partial charge of each O6 atom is negative.^[46] Dehydrated cations interact specifically with the guanine carbonyl oxygen O6 to stabilize the channel and the electrostatic repulsion between the cations is shielded by the partial negative charges on the O6 oxygen. Although there could be empty sites between the quartets, these vacancies must have a very short lifetime (few ns, on the basis of MD simulations^[47]). It is worth to note, that Sherman and co-workers have recently synthesized the first water-soluble template-assembled synthetic G-quartets (TASQ) able to self-assemble in single G-quartet using a template platform in absence of cations.^[48] Since the template bears four guanosine derivatives, the G-tetrad assembly is intramolecular and the entropic penalty associated with the self-assembly process is reduced, resulting in a thermodynamically stable G-quartet even in absence of cations.

A solid-state NMR study by Wang and Wu on 5'-GMP self-assemblies indicated the following stabilization capability: $K^+ > NH_4^+ > Rb^+ > Na^+ > Li^+$.^[49] In particular, the stabilization effect of Li^+ is negligible. Tl^+ has been also reported to bind to G-quadruplex.^[50,51] Amongst the divalent metal cations, Sr^{2+} ,^[52-54] Ba^{2+} ^[55] and Pb^{2+} ,^[56,57] are the most effective in inducing G-quadruplex formation, although also Ca^{2+} has been crystallized together with Na^+ in the channel of the G-quadruplex formed by $[d(TGGGGT)]_4$ (PDB: 2GW0).^[58,59] Interestingly, Ca^{2+} , Co^{2+} , Mn^{2+} , Zn^{2+} , Ni^{2+} and Mg^{2+} are supposed to induce dissociation of the quadruplex by entering transiently open G-quartets and coordinating the guanine hydrogen bond acceptors N7 and O6, thus preventing the re-establishment of the Hoogsteen interaction.^[60]

According to Hardin and co-workers, K^+ , Na^+ , NH_4^+ , Sr^{2+} and Pb^{2+} are the most effective stabilizers.^[60]

Details of the cation coordination geometry have been mostly provided by X-ray crystallography. A selection of G-quadruplex structures crystallized with different mono- and divalent cations deposited in the PDB is displayed in **Figure 9** and **Figure 10**, while **Table 1** summarizes the geometric parameters characteristic of the ionic channel in each structure. Na^+ can be coordinated either in the G-tetrad plane with a planar coordination geometry or in between the successive G-tetrad planes in a bipyramidal antiprismatic geometry. In general, for Na^+ intermediate coordination geometries are allowed in order to minimize the cation electrostatic repulsion (see for example the ionic channel in PDB 352D, **Figure 9**).^[61] All the other metal cations are too large to fit in-plane and are thus coordinated in the bipyramidal antiprismatic geometry. The coordination of the Na^+ and K^+ is characterized best because of their physiological relevance.

G-quadruplex structures

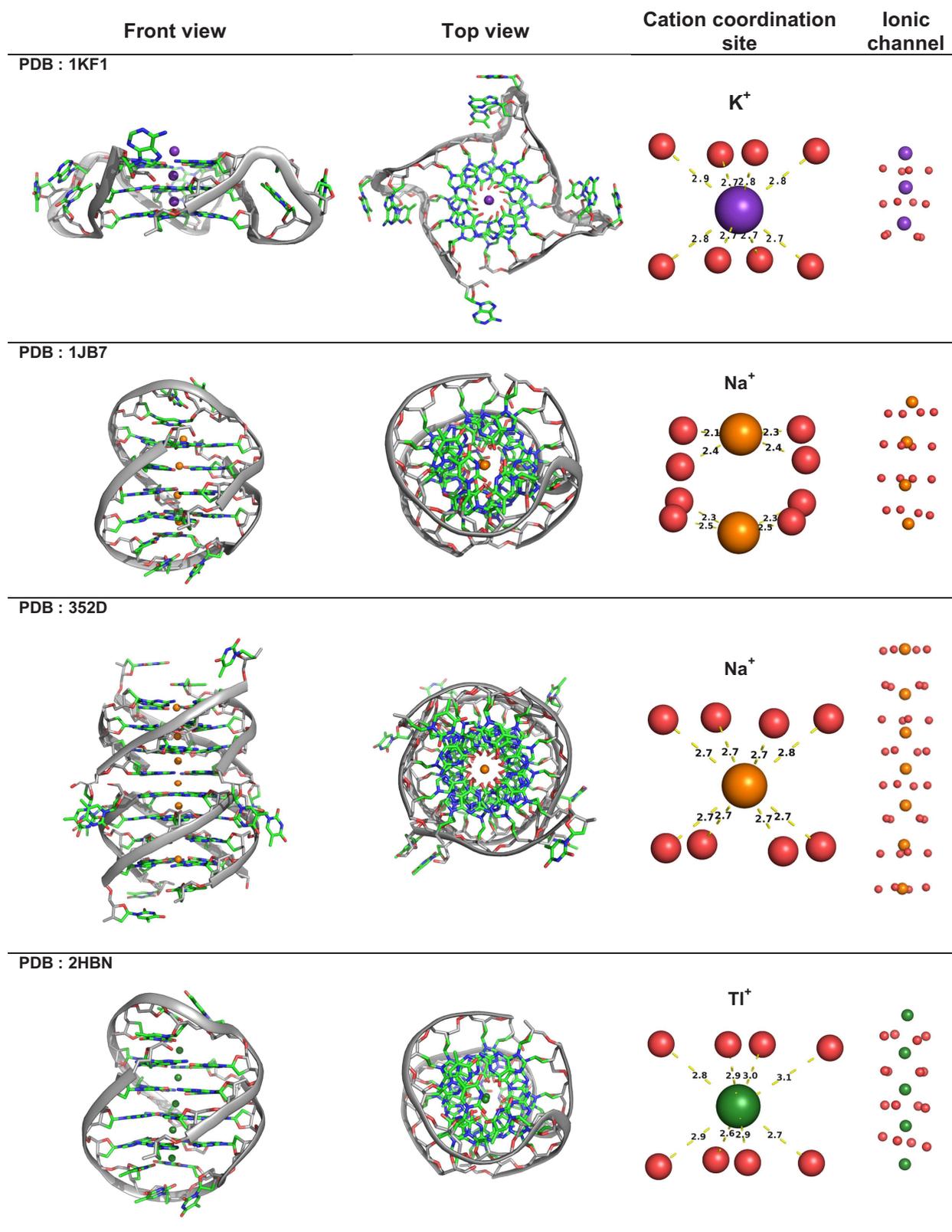


Figure 9 Coordination of monovalent cations in selected G-quadruplex crystal structures from the PDB. K⁺, Na⁺ and Tl⁺ cations are shown as violet, orange and green spheres, respectively. Details of the cation coordination site(s) and ionic channels formed by the O6 guanine atom (red spheres) are shown on the right. The distances are in Å.

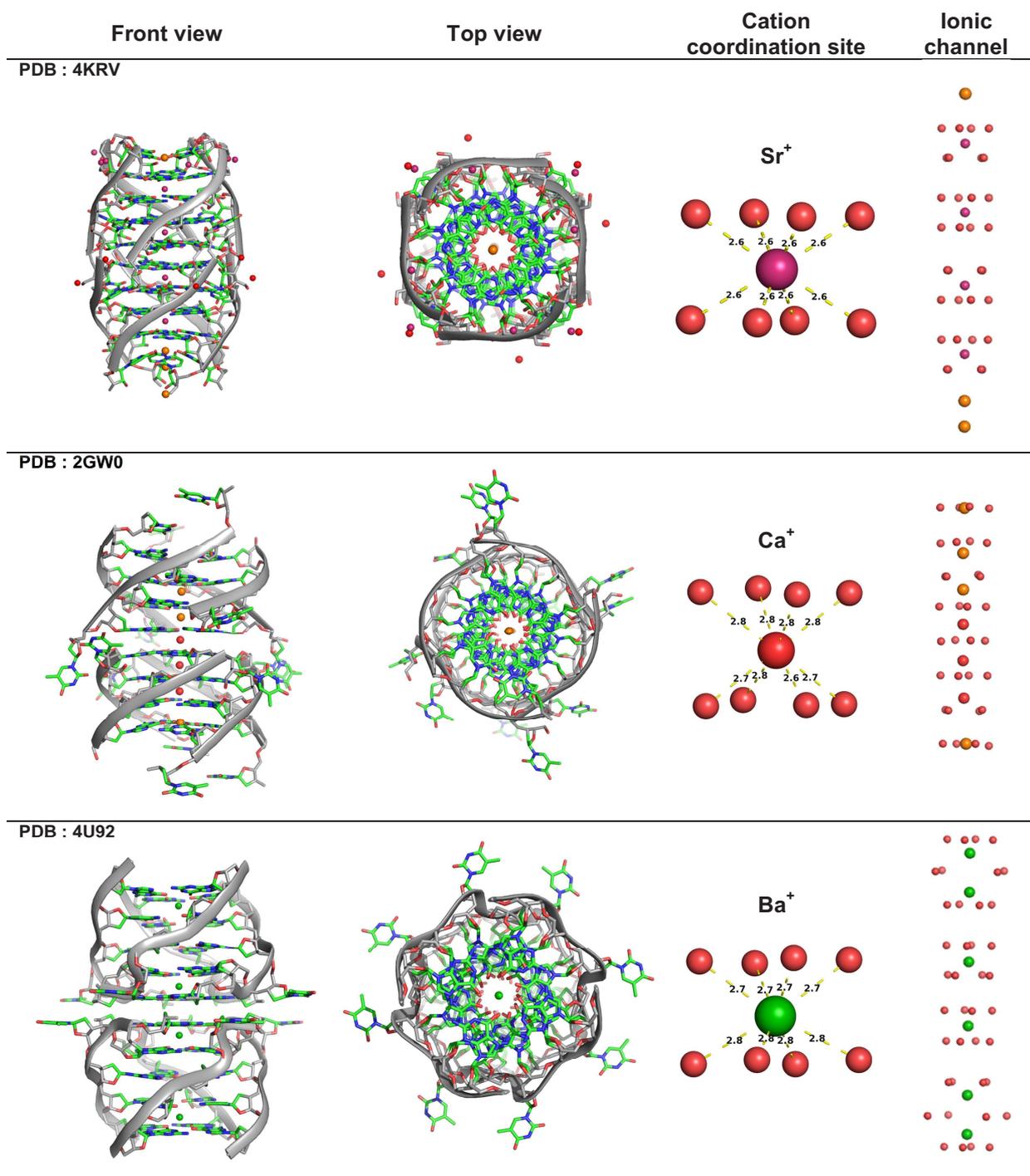


Figure 10 Coordination of divalent cations in selected G-quadruplex crystal structures from the PDB. Sr⁺, Ca⁺ and Ba⁺ cations are shown as magenta, red and green spheres, respectively. Structures 4KRV and 2GW0 contain additional Na⁺ cations (orange spheres) coordinated in the ionic channel. Details of the cation coordination site and ionic channels formed by the O6 guanine atom (red spheres) are shown on the right. The distances are in Å.

Cation	Effective ionic radius (Å)	Crystal radius (Å)	PDB ID (Res. (Å))	Sequence	Coordination distances (Å)	
K ⁺	1.51	1.65	1KF1 (2.10)	d[AGGG(TTAGGG) ₃]	2.4-3.2	
Na ⁺	0.99 (IV)	1.13 (IV)	1JB7 (1.86)	[d(GGGGTTTTGGGG)] ₂	2.1-2.5	2.3-2.7
	1.18 (VIII)	1.32 (VIII)	352D (0.95)	[d(TGGGGT)] ₄	2.3-2.4	2.5-2.9
Tl ⁺ (+Na ⁺)	1.59	1.73	2HBN (1.55)	[d(GGGGTTTTGGGG)] ₂	2.5-3.1	
Sr ²⁺	1.26	1.4	4RKV (0.88)	[d(UGGGGU)] ₄	2.6	
Ca ²⁺	1.12	1.26	2GW0 (1.55)	[d(TGGGGT)] ₄	2.5-3.1	
Ba ²⁺	1.42	1.56	4U92 (1.50)	[d(CCA ^{CNV} KGCGTGG)] ₄	2.7-2.8	

Table 1 Effective ionic radii and crystal radii of quadruplex stabilizing cations are reported according to Shannon.^[62] PDB identity, DNA sequence and coordination distances refer to the crystal structures shown in **Figure 9** and **Figure 10**.

The issues related to the determination of thermodynamic and kinetic parameters for the cation-induced dissociation and association of G-quadruplexes have been surveyed by Hardin and co-workers and Chaires and co-workers.^[60,63] Notably, solution studies show that, apart from the specifically bound K⁺ cations in the channel, there are more cations (either specifically or non-specifically/diffusely bound) than those detected in X-ray structures. The number of bound cations depends highly on the specific conformation(s) adopted by the G-quadruplex, which are extremely variable especially for telomeric G-quadruplex (see Chapter 2.3).^[64] Different distributions of populated conformations, resulting from differences in folding protocol, oligonucleotide sequence and experimental conditions (particularly the presence of fluorescent labels) and frequent misuse of a two-state transition model for interpretation of calorimetric and optical data partially explain the great variability reported in literature for the thermodynamics of cations binding to G-quadruplex.^[63]

G-quadruplexes coordinated to K⁺ are generally more stable than to Na⁺, as revealed by the melting temperatures (T_m). The higher thermodynamic stability of the K⁺-coordinated form is the result of two distinct contributions: the coordination energy (which is correlated to the size of the cation) and the cation dehydration energy. Thermodynamic analysis of competition experiments showed that the latter contribution is the actual driving force of the K⁺ over Na⁺ selectivity.^[65] Dingley *et al.* investigated by NMR the effect of cation and temperature on the G-tetrads hydrogen bond geometry on the sequence [d(G₄T₄G₄)₂], derived from *Oxytricha nova* telomeric DNA. The Na⁺ cations, located in the center of the tetrads, induce a tightening of the hydrogen bonding pattern, compared to K⁺ and NH₄⁺ cations. Furthermore, a global expansion of the tetrads is observed when increasing the temperature.^[66]

The guanine O6 atoms define a channel through which the cations can move along the coordination sites in the channel axis and not directly from the bulk to the channel or *vice versa*. This mechanism suggests the possibility of vacant sites, even if only with short lifetimes, when compared to the residence time of the cations in the coordination sites.

NH_4^+ has been used as probe for investigation of the movement of cations within the G-quadruplex channel by NMR (see Chapter 5.4). Feigon and co-workers have determined an exchange rate from the inner to the outer sites of 4 s^{-1} at 283 K for the symmetric, bimolecular, diagonally looped G-quadruplex formed by the sequence derived from *Oxytricha nova*. Addition of Na^+ cations to the $[\text{d}(\text{G}_4\text{T}_4\text{G}_4)]_2$ quadruplex pre-folded in presence of NH_4^+ cations results in a faster rate of NH_4^+ exchange from the inner to the outer sites (residence time decreased from 250 ms to 36 ms).^[67] These data further support the hypothesis that cations enter and exit from the external G-tetrads without disrupting the G-quartets, as the bound lifetime of the cations in the channel (hundreds of ms) is much shorter than the base-pairs opening rate (from hours to days). Analysis of the NH_4^+ cation coordination in the unimolecular quadruplex $\text{d}[\text{G}_4(\text{T}_4\text{G}_4)_3]$ revealed that the exchange rate of NH_4^+ between inner coordination site and the outer coordination sites is much slower ($0.05\text{-}0.07 \text{ s}^{-1}$) and that the exchange rate between outer sites $[\text{d}(\text{G}_4\text{T}_4\text{G}_4)]_2$ and the bulk is always faster than the exchange between inner and outer. Movement of cations from an outer binding site to the bulk is energetically less demanding than between inner sites. The cation movement within inner sites is hardly detectable for three tetrad G-quadruplexes. Moreover, the exchange rates reflect the rigidity of the quartets, since they have to slightly open in order to allow the movement of NH_4^+ .^[68] Therefore different loop geometries and possible stacking of loop residues may influence the external G-quartet stiffness and, ultimately, the cation exchange rate.^[69-72] Sket *et al.* have recently suggested that the presence of an all *syn* external tetrad drastically decelerates the cation movement into the bulk solution.^[73]

The nature of the stabilizing cation is a source of polymorphism. Different cations can induce diverse G-quadruplex structures, most notably for human telomeric DNA sequences (see Chapter 5.4 for details). A G-quadruplex conformational switch driven by the concentration ratio $[\text{Na}^+]:[\text{K}^+]$ has been reported first by Sen and Gilbert, who suggested a possible functional role for this switch,^[74] given the remarkable difference in average intracellular concentration of Na^+ (12 mM) *versus* K^+ (140 mM).^[75] Possible mechanisms for the interconversion between the Na^+ - and the K^+ -form of human telomeric sequences have been proposed, on the basis of CD, 2-aminopurine fluorescence and NMR data.^[76-78] Physiological conditions are characterized by a complex mixture of mono- and divalent cations; therefore the existence of a G-quadruplex coordinating different cations has been hypothesized. In fact, a form containing a 1:1 mixture of $[\text{NH}_4^+]:[\text{K}^+]$ has been identified by NMR for the quadruplex formed by $[\text{d}(\text{G}_3\text{T}_4\text{G}_4)]_2$.^[79] Moreover, a tetrameric quadruplex containing a mixture of Ti^+ and Na^+ has been crystallized (PDB code: 1S45).^[80] On the other side, there are also G-quadruplex forming sequences that conserve the same folding topology in presence of different stabilizing cations. For example, it was shown that $[\text{d}(\text{G}_4\text{T}_4\text{G}_4)]_2$ folds into an anti-parallel, dimeric, diagonally looped structure in the presence of Na^+ , K^+ and NH_4^+ . There are, however, minor differences between the T_4 loops conformations which can be attributed to different cation coordination modes. In fact, while the size of K^+ and NH_4^+ allows only inter-plane coordination, Na^+ can be coordinated within the G-tetrad plane and can additionally coordinate the O2 atom of a thymine.^[81] It has recently been observed that the 23 nt-long $\text{d}[\text{TAGGG}(\text{TTAGGG})_3]$ derived from the human telomeric sequence adopts the same folding topology (hybrid-1) in the

presence of K^+ and in the presence of Na^+ . The K^+ -form is more stable ($\Delta T_m = 10$ K) and addition of K^+ cations readily replaces the Na^+ cations in the G-quadruplex core in a stepwise process that does not require global unfolding.^[82]

2.2 Stacking of G-tetrads

The geometry of stacking is commonly defined according to the polarity of consecutive tetrads and to the overlap of the stacked bases.

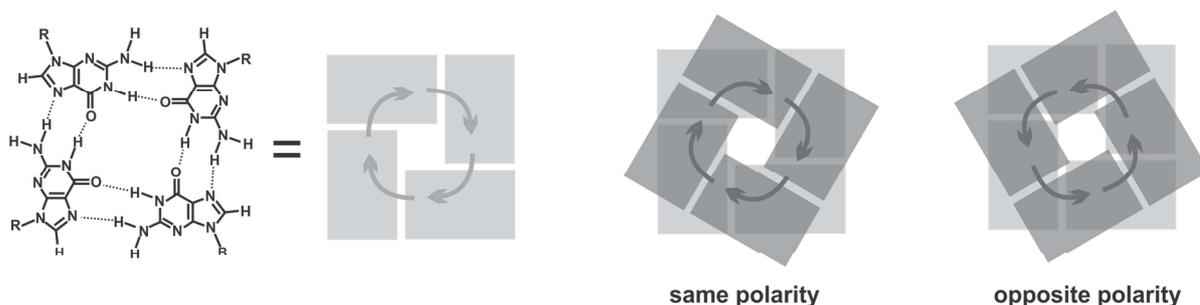


Figure 11 Definition of stacking polarity in two consecutive G-quartets.

The polarity of Hoogsteen hydrogen bonds within a G-tetrad can be defined following the donor-acceptor pattern as shown in **Figure 11**. Two consecutive G-quartets can have same polarity or opposite polarity of the Hoogsteen hydrogen bonding pattern. It has recently been shown that the flipping of the hydrogen bond polarity can be achieved within a tetrad by selective substitution of *syn* guanine residues with the *anti*-favoring 2'-deoxy-2'-fluoro-guanosine (2'-F-r)^[83] or by rational incorporation of 8-oxo-guanine and xanthine in a complementary fashion.^[84]

A survey of structures determined by X-ray crystallography was examined by Lech *et al.* to identify the most commonly observed stacking geometries.^[85] The statistical analysis of the database revealed that in the G-quadruplex core there are three recurring stacking geometries (**Figure 12**). The 5-5 geometry (**Figure 12a**) is characterized by an overlap of the 5-membered rings at 5'-*syn-anti-3'* steps in tetrads with opposite polarity. The 5-6 geometry (**Figure 12b**) features a partial overlap of the 5-membered ring of one tetrad with the 6-membered ring of the other one and occurs at 5'-*anti-anti-3'* steps in tetrads with the same polarity. A partial overlap of the 6-membered rings (6-6 geometry in **Figure 12c**) is observed at 5'-*anti-syn-3'* steps in G-tetrads with opposite polarity. Thus, a strong correlation between glycosidic torsion angle and stacking base is observed. The statistical analysis of the crystallographic database gave similar results to those obtained on a selection of NMR structures. Furthermore, a folding topology reported in solution but never observed in crystal structures features 5'-*syn-syn-3'* steps that adopt the same partial 5-6 stacking geometry observed at 5'-*anti-anti-3'* steps in crystal structures.^[85]

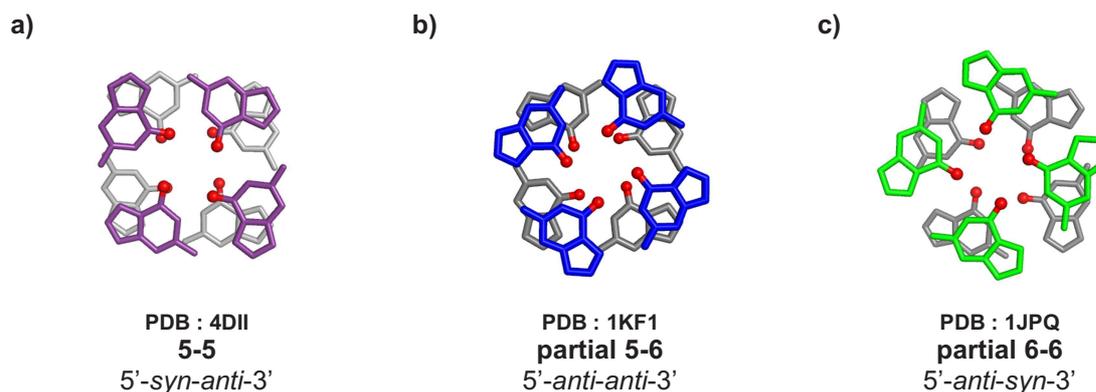


Figure 12 Examples of the most frequently observed stacking geometries, according to the database of X-ray crystal structures examined by Lech *et al.*^[85] The carbonyl O6 atoms in the guanine residues are indicated with red spheres.

Various other geometries of base stacking can be observed at the interface of stacked G-quadruplexes.^[85]

2.3 Polymorphism of G-quadruplex

G-quadruplex structures are extremely diverse and a variety of structural elements has been reported. A G-quadruplex can be formed inter-molecularly between four strands (tetramolecular) or between two strands (bimolecular), as well as intra-molecularly (unimolecular), as indicated in **Figure 13a**. The backbone strands that connect the G-stretches involved in the formation of the tetrads can assume different relative 5' to 3' orientations (indicated with arrows in selected examples on top of **Figure 13b**). The G-quadruplexes can feature four strands with the same orientation (all parallel), three strands oriented in one direction and one in the other (hybrid 3+1 topology) or a mixture of two strands in one orientation and two in the other (anti-parallel 2+2 topology), with different possibilities of relative orientation (see for example PDB 4DII for $\uparrow\downarrow\uparrow\downarrow$ geometry and PDB 143D for $\uparrow\uparrow\downarrow\downarrow$ geometry). The backbone strands connecting guanine residues at adjacent positions in contiguous tetrads are usually continuous, but they can also present interruptions, as shown for example in the PDB structure 2KPR. The loops are structural elements connecting two G-stretches of the G-quadruplex core and can adopt four different geometries, indicated in different colors on the selected unimolecular quadruplex structures displayed in **Figure 13b**. The double-chain reversal loop (orange color code) connects two guanine bases belonging to different tetrads in adjacent, parallel strands. Guanine bases sharing hydrogen bonds and belonging to antiparallel strands are connected by edgewise loops (red color code), while diagonal loops connect opposite guanine residues in the same tetrad, belonging to non-adjacent, anti-parallel strands (cyan color code).

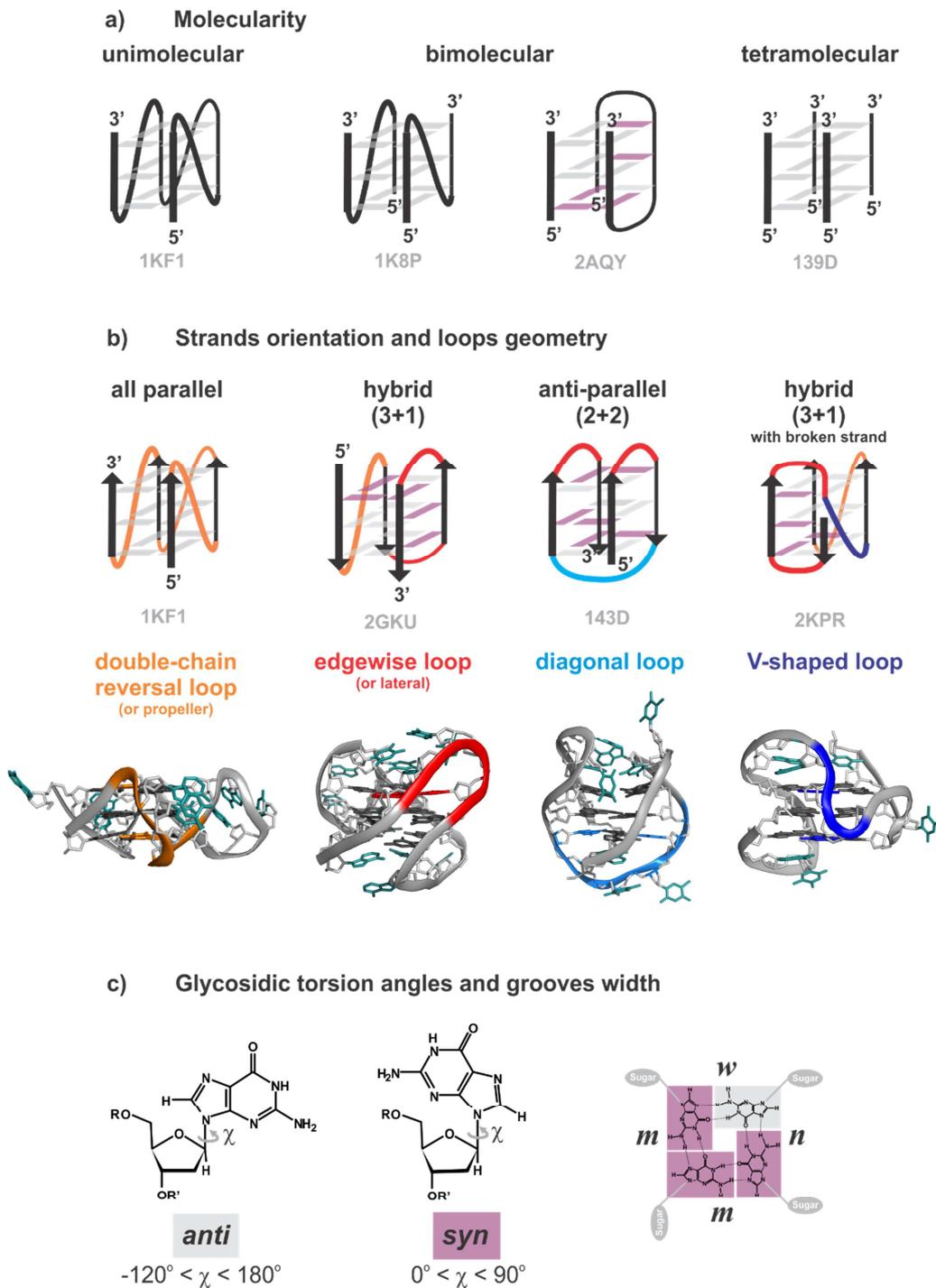


Figure 13 Structural elements defining G-quadruplex polymorphism: a) molecularity, b) strands orientation and loops geometry, c) glycosidic torsion angles and grooves width. The PDB codes of the selected examples are indicated in gray below each structure. In the schemes displayed on top of panel b) double-chain reversal loops are colored in orange, edgewise loops in red, diagonal loops in cyan and V-shaped loops in blue. In the bottom of panel b) a double-chain reversal loop is highlighted in orange on PDB structure 1KF1, an edgewise loop is highlighted in red on PDB structure 2GKU, a diagonal loop is highlighted in cyan on PDB structure 143D and a V-shaped loop is highlighted in blue on the PDB structure 2KPR. On each PDB structure, backbone and sugar are light gray, loops residues are marine green, G-tetrad's guanines connected by the highlighted loop are colored with the same color code used to highlight the loop, while the others are dark gray. Guanine residues in *anti* and *syn* conformation are indicated in gray and purple, respectively. The groove width is indicated as example on the G-tetrad on the right of panel c). *w*, *m* and *n* stand for wide, medium and narrow groove, respectively.

V-shaped loops were observed for the first time in 2001 by Patel and co-workers and connect two corners of a G-tetrad core lacking one supporting column (blue color code in **Figure 13b**).^[86] It was suggested that the break in the backbone could facilitate the insertion of a planar ligand between the G-tetrad planes. However, no intercalating G-quadruplex ligand has been reported up to now.

The length and the composition of the loops are important factors in determining topology and stability of the quadruplex. Several groups have systematically investigated the influence of the loop length on 3-tetrads intramolecular quadruplexes.^[87,88] The ultimate aim of these studies is to find a rationale for predicting the thermodynamic stability and the folding topology of G-quadruplex solely on the base of the nucleotide sequence. The longer the loop, the less stable is the quadruplex, on the basis of the melting temperature (T_m) measured by UV or CD. One long loop (more than 7 nt-long) can be tolerated, but incorporation of two or three long loops is destructive.^[88] While V-shaped loops contain no nucleotides^[86,89-91] and diagonal loops usually require at least 3 nucleotides, single-nucleotide loops favor the formation of double-chain reversal loop, which is a very robust scaffold in G-quadruplexes from gene promoter regions. Indeed, most of the quadruplexes formed in the gene promoter regions contain two double-chain reversal 1 nt loops and one central double-chain reversal loop of variable length (see **Table 5**). Interestingly, a very stable all-parallel quadruplex containing a 13 nt-long central loop has been recently characterized by Yang and co-workers in the human BCL-2 proximal promoter region.^[92] Webba da Silva and Plavec showed that it is possible to design the oligonucleotide sequence in order to control the G-quadruplex self-assembly and guide the folding into a specific topology just by controlling the length of the loops.^[93] The design is successful for G-quadruplexes with loops containing only thymine residues, which are less prone to hydrogen bonding and stacking. In fact, the composition of the loop plays also a role. Loop residues cannot only form additional base pairs, but also stack on top of each other or on top of an external G-quartet, conferring to the quadruplex additional stability. It has been reported that adenine residues belonging to single nucleotide loops in 2-tetrad quadruplexes can arrange along the edges of the G-tetrad *via* hydrogen bonds to form a mixed G/A higher order platform, such as an heptad.^[94] The different thermodynamic stability of two G-quadruplex structures formed in the nuclease hypersensitivity element III₁ (NHE III₁) of the *c-MYC* gene (see Chapter 3.2) was also explained in terms of loop composition: the more stable quadruplex contains two double chain reversal single-thymine loops while the less stable one contains two double chain reversal single-adenine loops. In a double chain reversal single-nucleotide loop, the adenine adopts a “wing-up” conformation, with the base almost perpendicular to the groove, which is more solvent exposed than the “wing-down” conformation adopted by a thymine, which is reported to lay in the groove. Adenine is more solvent-protruding than thymine, therefore it is suggested to be energetically less favored.^[95] The analysis of the effect of the sequence composition on a 3 nt-long central loop showed also that an adenine residue at the 5'-end of the loop is destabilizing.^[96]

The structural diversity is furthermore determined by the conformation of the glycosidic bond angle (χ , **Figure 13c**), which can assume either *anti* or *syn* orientation. The glycosidic bond angles of two adjacent bases define the width of the groove (**Figure 13c**, right): narrow (n) and wide (w) grooves result from contiguous bases with *anti-syn* or *syn-anti* conformation, while medium (m) grooves are defined by contiguous bases with *anti-anti* or *syn-syn* conformation. Thus, in a tetrad there are 16 possible combinations of glycosidic bond angles, which define 8 possible groove width combinations. Only G-quartet characterized by the same groove width combination can stack on top of each other to form a stable quadruplex. An accurate geometrical formalism for description of G-quadruplex folding topologies has been developed by Webba da Silva in 2007.^[97] **Figure 14** displays as example three different grooves observed in the hybrid (3+1) structure formed by the telomeric sequence Tel24 d[TTGGG(TTAGGG)₃A].

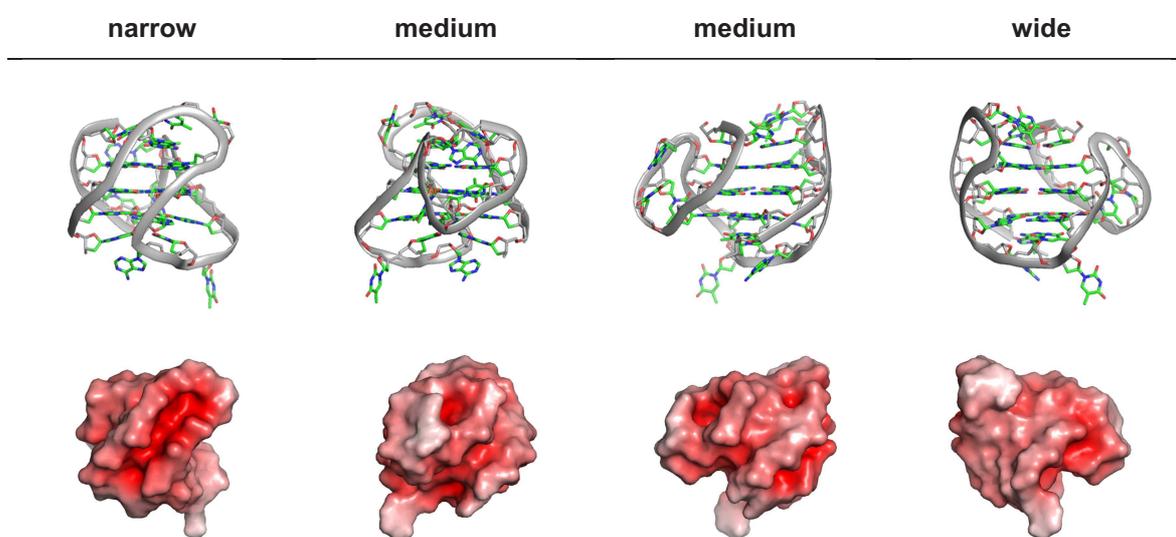


Figure 14 Examples of different grooves for the hybrid (3+1) quadruplex formed by the telomeric sequence Tel24 d[TTGGG(TTAGGG)₃A] (PDB: 2GKU). Narrow, medium and wide grooves are highlighted on the NMR structure (top) and on the corresponding solvent accessible surface area (SASA) colored by electrostatic potential (bottom), calculated using the Adaptive Poisson-Boltzmann Solver (APBS) Tool implemented in Pymol.

Interestingly, the first structure of a left-handed G-quadruplex has just recently been reported by Phan and co-workers (PDB: 2MS9, NMR structure; PDB: 4U5M, X-ray structure). The CD signature of the “Z-G-quadruplex” is peculiar and the backbone dihedrals are similar to those reported for Z-DNA. The stacking distances and the groove geometry are comparable with those of a right-handed G-quadruplex, but the progression of the backbone in two successive guanines is left-handed.^[98]

2.4 Polymorphism of human telomeric DNA G-quadruplex

A well-studied example of G-quadruplex polymorphism is given by the variety of structures adopted by the sequences derived from the four G-repeats d[GGG(TTAGGG)₃], which is the minimal motif in human telomeric DNA able to form a unimolecular G-quadruplex (See Chapter 3.1).^[99,100] Besides NMR and X-ray, a plethora of experimental methods, including circular dichroism (CD),^[101,102] mass spectrometry,^[103] single molecule Förster resonance energy transfer (sm-FRET),^[104,105] electron paramagnetic resonance (EPR),^[106,107] fluorescence spectroscopy^[108,109] and calorimetry,^[110] has revealed the complex nature of G-quadruplex polymorphism in human telomeres.

The overall structure depends on the nature of the stabilizing cation, on the presence of molecular crowding agents (such as polyethylene glycol) and on the presence and composition of 5'- and 3'-flanking nucleotides adjacent to the d[GGG(TTAGGG)₃] core. The various topologies reported until now for sequences derived from d[GGG(TTAGGG)₃] are summarized and visualized in **Figure 15**, **Figure 16** and **Table 2**.

The first structure of unimolecular telomeric quadruplex was reported in 1993 for the sequence d[AGGG(TTAGGG)₃] in Na⁺-containing solution (PDB: 143D).^[111] It folds in the so-called “basket” topology, with three tetrads and anti-parallel strands connected by lateral-diagonal-lateral loops. However, in K⁺-containing solution this sequence forms a mixture of more multiple conformations.^[112] The X-ray crystal structure of the same DNA crystallized in presence of K⁺ cations was published in 2002 by Neidle and co-workers and was characterized by a completely different folding topology commonly indicated as “propeller” with all the guanine residues in *anti* conformation and all parallel strands connected by propeller loops (PDB: 1KF1).^[113] Interestingly, the propeller G-quadruplex is the only folding topology observed in crystal structures irrespective of the telomeric sequence used (see for example PDB structures: 3CDM 23 nt-long, 4DA3 21 nt-long, 1K8P bimolecular 12 nt-long). The propeller structure has been observed in solution only in presence of 40% polyethylene glycol 200 (PEG 200) (PDB: 2LD8),^[114] but it is still matter of debate, whether the telomeric quadruplex is capable of adopting this topology in solution. Two different structures (“hybrid-1” and “hybrid-2”) consisting of three stacked G-tetrads connected by a mixture of three parallel strands and one antiparallel strand (hybrid 3+1 forms) were identified in K⁺-containing solution by Yang group and Phan group.^[115–117] The two topologies differ in the successive order of the loop arrangement, which is propeller-lateral-lateral in the hybrid-1 and lateral-lateral-propeller in the hybrid-2. Noteworthy, the human telomeric sequence d[TAGGG(TTAGGG)₃] is able to adopt the hybrid-1 topology either in K⁺ or in Na⁺-containing solution.^[82]

A two-tetrads antiparallel basket structure has been also characterized in K⁺-containing solution with unmodified sequences lacking the 5'-flanking nucleotides.^[118,119] This two-tetrad quadruplex features an extensive loop-loop interaction *via* base pairing. In 2013 a novel antiparallel (2+2) topology with antiparallel strands connected by lateral-propeller-lateral loops has been

characterized by NMR in Na⁺-containing solution.^[120] Although there is no atomic resolution structure reported until now, the “chair” folding, featured by antiparallel strands, has been proposed as a plausible structure.^[121]

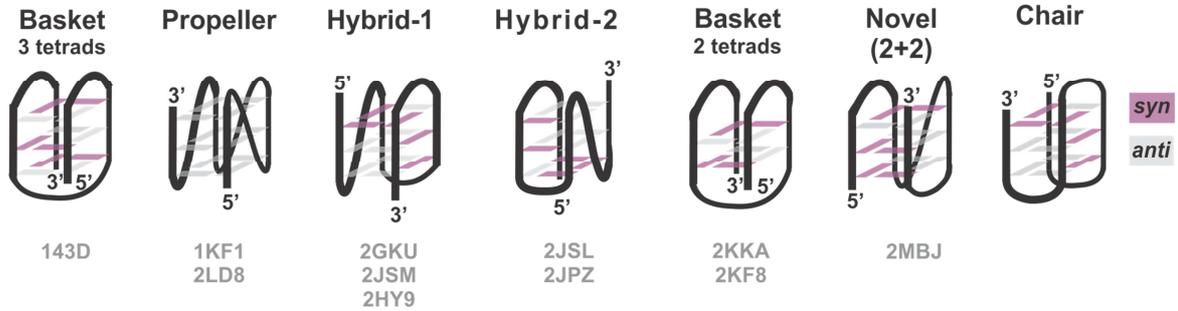


Figure 15 Schematic representation of the major folding topologies proposed for unimolecular human telomeric quadruplex. For each topology, the common name (top) and the PDB code (bottom, gray) of the available structures are indicated. Guanine residues in *anti* and *syn* conformation are shown in gray and purple, respectively.

G4 topology	Sequence	Cation	PDB code	Method	Topology
Basket 3-terads	d(AGGGTTAGGGTTAGGGTTAGGG)	Na ⁺	143D	NMR	(2+2), (l, d, l)
Propeller	d(AGGGTTAGGGTTAGGGTTAGGG)	K ⁺	1KF1	X-ray	All parallel, (p, p, p)
	d(TAGGGTTAGGGTTAGGGTTAGGG)	K ⁺ (*)	2LD8	NMR	
Hybrid-1	d(AAAGGGTTAGGGTTAGGGTTAGGGAA)	K ⁺	2HY9	NMR	(3+1), (p, l, l)
	d(TTAGGGTTAGGGTTAGGGTTAGGGGA)	K ⁺	2GKU	NMR	
Hybrid-2	d(TAGGGTTAGGGTTAGGGTTAGGGT)	K ⁺	2JSM	NMR	(3+1), (l, l, p)
	d(TTAGGGTTAGGGTTAGGGTTAGGGTT)	K ⁺	2JPZ	NMR	
Basket 2-tetrads	d(AGGGTTAGGGTTAGGGTTAGGGT)	K ⁺	2KKA	NMR	(2+2), (l, d, l)
	d(GGGTTAGGGTTAGGGTTAGGGT)	K ⁺	2KF8	NMR	
Novel (2+2)	d(TTAGGGTTAGGGTTAGGGTTA ^B rGGGTTA)	Na ⁺	2MBJ	NMR	(2+2), (l, p, l)

Table 2 Overview of the unimolecular G-quadruplex (G4) structures derived by human telomeric sequences deposited in the PDB up today. Mutations compared to the wild type are highlighted in red. (*) 40% PEG 200. The topology is described on the basis of the relative orientations of the strands (all parallel, 2+2 or 3+1) and of the geometry of the loops encountered in direction 5' to 3' (l, d and p indicate lateral, diagonal and propeller loop, respectively).

G4 topology	PDB code		Groove width (Å)		
			Narrow	Medium	Wide
Hybrid (3+1)	Form 1	2GKU	11.0 - 11.8	15.5 - 16.4	19.7 - 19.9
	Form 2	2JSL	10.8 - 11.1	16.4 - 17.0	19.1 - 18.4
Propeller	1XAV		-	15.3 - 18.1	-
Basket 3-tetrads	134D		11.2 - 13.6	13.7 - 16.8	16.0 - 18.1
Z-G-quadruplex	2MS9		-	14.3 - 17.4	-

Table 3 Groove width measured as phosphorous-phosphorous distance on adjacent guanine residues in the inner tetrad of selected G-quadruplex (G4) structures.

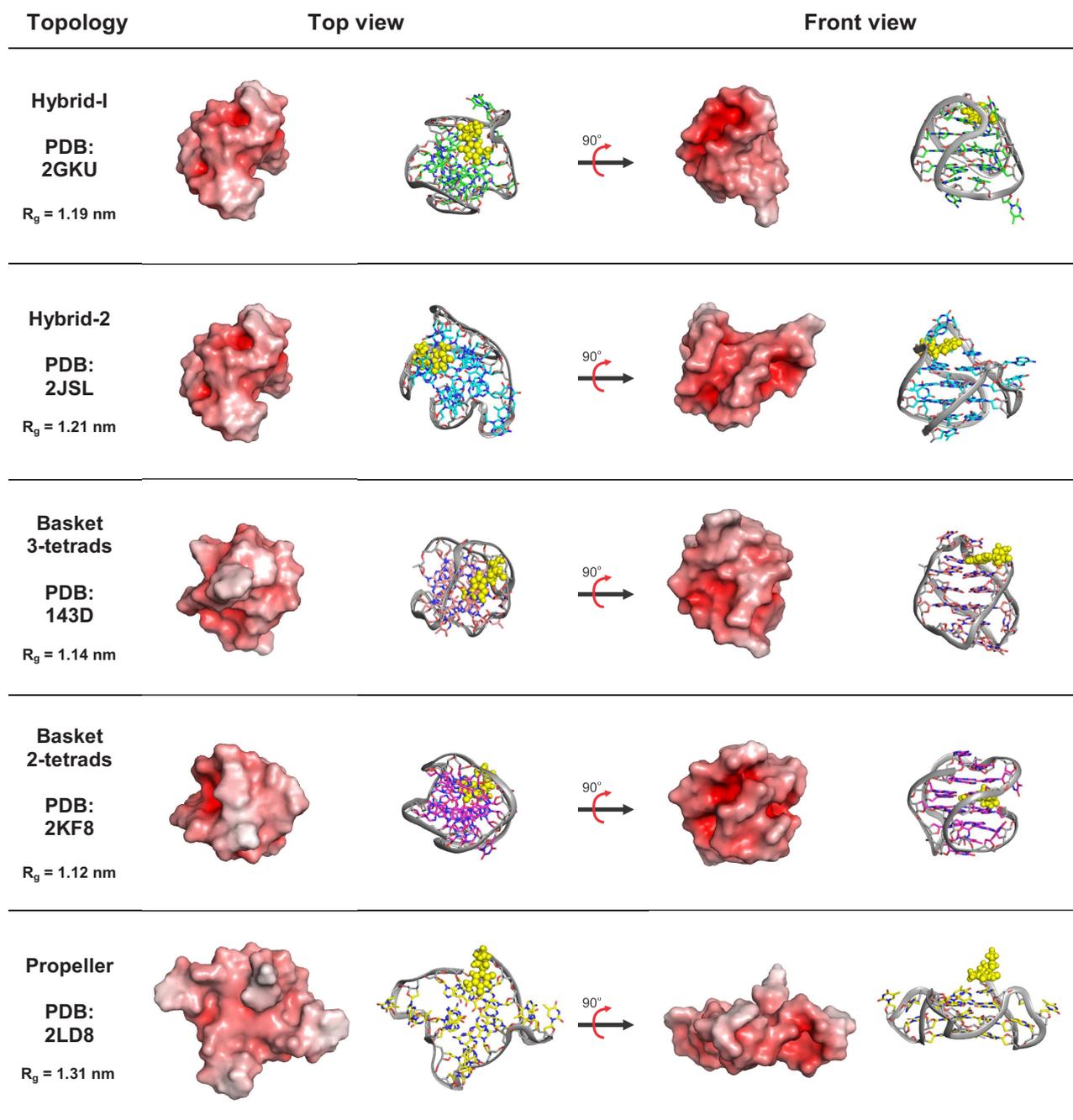


Figure 16 Top view (5'-end) and front view of representative NMR structures reported for human telomeric DNA. Structures are displayed as solvent accessible surface area (SASA) colored by electrostatic potential (left) and as sticks (right), with the 5'-end residue shown in yellow spheres. The SASA and the electrostatic potential were calculated using the Adaptive Poisson-Boltzmann Solver (APBS) Tool implemented in Pymol. The radius of gyration (R_g), calculated for the displayed structure with the software HYDROPRO,^[122] is indicated in the first column.

It is still “not so crystal clear”^[123] which of these conformations is adopted by the telomeric 3'-end overhang *in vivo*. Most of the structures presented above were determined in experimental conditions which are very far away from the structured and complex environment that the biomolecules experience *in cell*. In fact, in eukaryotic cells biomacromolecules occupy around the 30% of the total volume and small organic molecules that act as osmolytes can reach the molar

range, making excluded volume effects crucial in perturbing the conformational equilibria.^[124] Moreover, DNA in the nucleus is highly packed in nucleosomes and metaphase chromosomes and the single-stranded telomeric ends are engaged in interactions with several proteins, particularly with the shelterin protein complex (see section 3.1), which further complicates the development of appropriate *in vitro* models for mimicking the *in cell* conditions. Most of the experimental results reported until now are performed in diluted solution or in presence of molecular crowding reagents such as polyethylene glycol (PEG), whose effect on the G-quadruplex structures is not completely understood. Human telomeric DNA sequences have been shown to adopt the same propeller conformation as crystal and in solution containing 40% PEG 200.^[113,114] Since the effect of PEG was initially suggested to be molecular crowding, the propeller conformation was proposed to be the physiologically relevant one.^[114] However, this is still a matter of debate in the G-quadruplex community.^[114,125–128] The capability of 40% PEG 200 to induce the conversion of any human telomeric quadruplex conformation to the all-parallel stranded topology was already reported in 2007,^[129] and exploited by Chang and co-workers in 2012 to engineer carbazole ligands with polyethylene glycol units and to induce conformational changes upon binding.^[130] Other molecular crowding agents like Ficoll 400, bovine serum albumin (BSA) and *Xenopus laevis* egg extract are not capable to induce any conformational transition in telomeric G-quadruplex.^[125,126] Molecular crowding acts on the basis of shape and size differences, therefore, for molecular crowding to be the determinant mechanism to stabilize the propeller conformation, the propeller form should be the most compact. The radius of gyration characteristic of different topologies and the solvent accessible surface (see **Figure 16**) show, however, that the propeller form is not the most compact conformation.^[125] Therefore, molecular crowding is not the major determinant of the PEG-driven transition to the propeller form. Hydration effects are distinct from crowding and rely on the change in water activity. While the formation of duplex and triplex DNA is accompanied by an increase in water activity (the folded structures are more hydrated than the single-stranded), the formation of G-quadruplexes is associated with a release of water molecules and is favoured by a reduction in water activity. Some authors explained the conformational transition usually observed in presence of PEG in terms of dehydration effects.^[125] Another possible explanation suggested is that PEG binds preferentially to the propeller form with a conformational selection mechanism.^[127] The discussion is still open, but it is meanwhile accepted that PEG is not the most appropriate crowding agent to investigate the effect of molecular crowding on telomeric DNA.^[126]

It is important to keep in mind that most of the studies are performed with 4 telomeric repeats, which are basically just a very shortened version of the 150-250 nt-long single stranded DNA present at the 3'-end of telomeres. It has been proposed that *in vivo* more G-quadruplex units could be formed at the telomeres. In fact, there is experimental evidence for the formation of higher order G-quadruplex structures in long (8-16 repeats) telomeric DNA sequences.^[101,131,132] However, the details of the structural arrangement of the human telomeric overhang are still controversial.^[133] In particular, there is no consensus regarding the interaction mode between adjacent quadruplex units. Single-molecule laser tweezer experiments conducted on 12-24 repeats of telomeric sequences show that the G-quadruplexes are likely non-interacting, following

the so-called “beads-on-a-string” model (**Figure 17a**).^[134] On the other hand, molecular dynamics (MD) simulations supported by sedimentation and fluorescence studies suggest that the adjacent G-quadruplex units are interacting via the loops and can therefore not be considered completely independent (**Figure 17b**).^[132,135] The folding topology of the G-quadruplex units involved in higher order structures is also not clear. The packing of propeller (P in **Figure 17b**) units via G-tetrads interaction is favoured in presence of 40% PEG 200,^[136] but can be excluded on the basis of sedimentation coefficient data, which indicate an hybrid-based (H1 and H2 in **Figure 17b**) model.^[132] Furthermore, NMR studies have recently provided an atomic resolution insight into the folding of a telomeric sequence with 8 repeats, suggesting that *ex vivo* (*in cell extract*) it might adopt both 2-tetrad basket and hybrid-2 conformations.^[137]

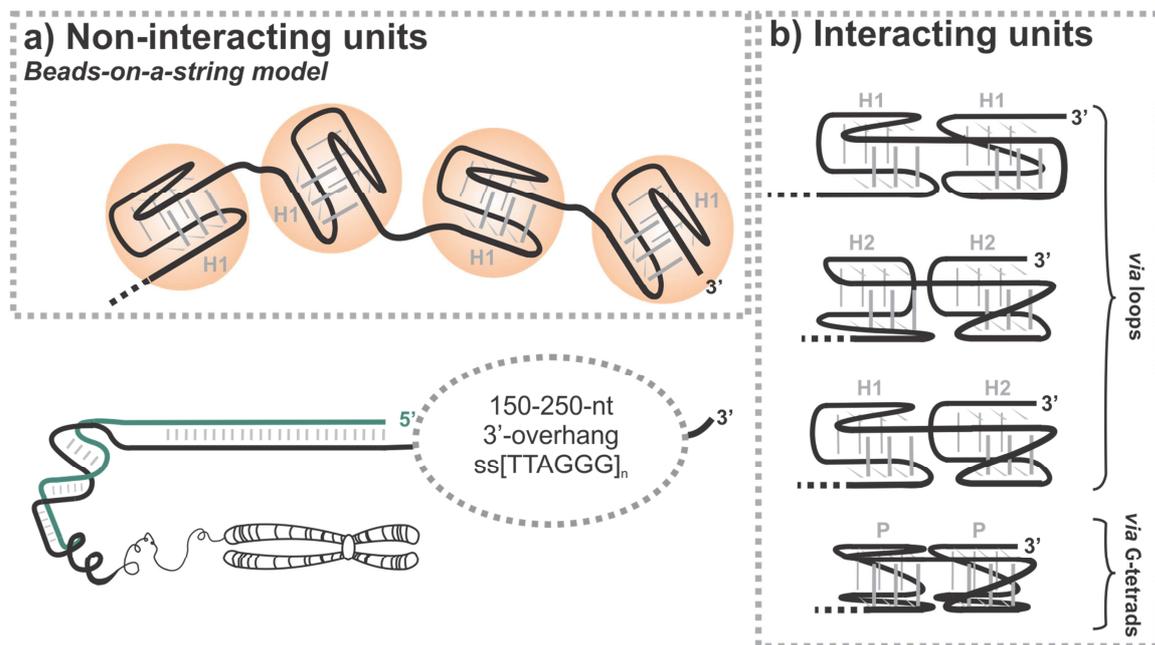


Figure 17 Proposed models for the arrangement of G-quadruplex units in long human telomeric 3'-overhang. a) Non-interacting units (beads-on-a-string model). b) Interacting units via G-tetrads stacking of propeller G-quadruplex units (P) or via loop residues interaction of hybrid-1 (H1) and hybrid-2 (H2) G-quadruplex units.

3 Location of G-quadruplexes in the genome and biological function

Computational analyses showed that potential G-quadruplex forming motifs are non-randomly distributed and evolutionary conserved in the human genome, suggesting that G-quadruplex structures may play key roles in a number of biological processes. Maizels *et al.* has recently introduced the concept of “G-quadruplex genome”, as the set of G-quadruplex motifs that participate in key processes.^[138]

The first reports on the total number of potential quadruplex-forming sequences in the human genome were published one decade ago.^[139–141] Over 350,000 putative quadruplex sequences (PQS) were identified with two similar approaches,^[139,140] based on the following pattern-based “Folding Rule”:



with n being the number of G stacks and l_1 , l_2 , l_3 being the number of nucleotides belonging to loop 1, 2 and 3, respectively. The restraints for the parameters n , l_1 , l_2 and l_3 vary in different algorithms. The maximum length of the loop is usually set to 7 nucleotides, on the basis of molecular modelling^[142] and biophysical studies^[87] suggesting that an increase in the loop length has a major effect on the G-quadruplex stability. However, Yang and co-workers have recently shown that the major G-quadruplex conformation formed by the purine rich sequence involved in the regulation of the gene *BCL-2* (see Chapter 3.2) contains a 13-nt long chain reversal loop,^[92] revealing the necessity to refine the prediction algorithm. Any prediction algorithm relying on a pattern-based folding rule like the one discussed above does not predict the thermodynamic stability of the G-quadruplex. More recently, Stegle *et al.* have developed an algorithm to infer the G-quadruplex thermal stability from the sequence.^[143] Furthermore, databases for predicting G-quadruplex formation and stability are available online.^[144,145] Despite these limitations, bioinformatic approaches have led to many discoveries.^[146]

A different computational approach was adopted by Maizels *et al.* to quantify the potential of genomic regions to form G-quadruplexes (G4P).^[147] The “G4P calculator” software (<http://depts.washington.edu/maizels9/G4calc.php>) evaluates whether sliding windows of 100 bases in length contain at least four runs with three or more guanines, independently on their separation.

PQS are relatively rare in human genome: approximately a third less than what would be expected by chance, suggesting evolutionary selective pressure against G-quadruplex formation in mRNA.^[139] However, PQS are highly abundant in gene promoters, defined as the region 1 kb upstream of the transcription start site (TSS), compared to the genome average.^[148] In particular,

G-quadruplex are highly concentrated few hundreds bases immediately upstream of the TSS and are supposed to be more stable than in the rest of the genome, because they are rich in one-nucleotide loops.^[143] More than 40% of the promoter regions of annotated genes contain one or more PQS. It is also striking that DNase I hypersensitive sites (NHS)^[149] are enriched in PQS compared to the average genome, implying that these regions undergo a structural transition from double helix DNA and an unwound form, which is a prerequisite in gene promoters for transcription activation. Furthermore, binding to storage proteins like histones would impair the formation of G-quadruplexes. Indeed, G-quadruplexes are likely to be formed outside the nucleosome.^[150] There is a close correlation between G4P in template and non-template strand. This potential is however low in the non-template strand, which is connected to the formation of transcription-induced structures.

Maizels *et al.* did not only enumerate the regions with G4P, but also showed that there is a correlation between G4P and gene functions. G-protein-coupled receptors, sensory perception, nucleosome assembly, ubiquitination and nucleic acid binding are examples of functions characterized by low G4P, while functions such as transcription factor activity, cell signalling, muscle contraction, development, growth factors and cytokines present high G4P. Interestingly, the distribution of G4P for tumor suppressor genes was shifted from the RefSeq (RefSeq, <http://www.ncbi.nlm.nih.gov/refseq/rsg/about/>) median towards lower values, while the distribution of proto-oncogenes was shifted towards higher G4P. Therefore, tumor suppressors and proto-oncogenes are characterized by low and high G4P, respectively.

Analysis of the G4P frequency in genomic sequences used as reference standards for well-characterized genes (**Figure 18**) shows that motifs with high potential of forming G-quadruplexes (with loop size 1-12 nt) are frequent at the transcription start site (TSS), at the 5'-UTR and at the 5'-end of the first intron, but most of the coding regions are depleted of potential G-quadruplex forming motifs.

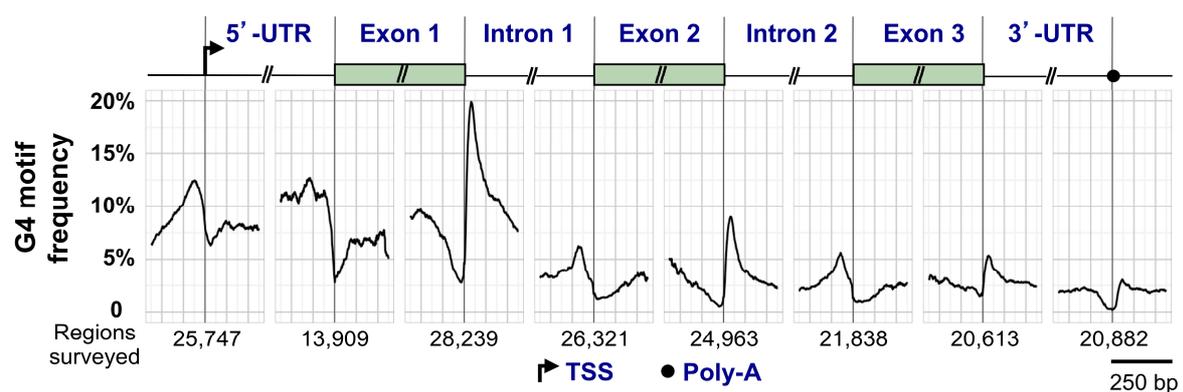


Figure 18 Distribution of regions with high potential to form G-quadruplex (G4) in a generic human RefSeq gene. Extracted from Reference [138]. The transcription start site (TSS) is indicated with an arrow.

Although bioinformatics pointed out that functional regions of genome have a high potential for G-quadruplex formation, the existence of G-quadruplex structures *in vivo* has been matter of debate.^[151] In particular, strong evidence for the formation of G-quadruplexes at telomeres in

humans was reported only very recently from the Balasubramanian group using G-quadruplex specific engineered antibodies.^[152] Engineered antibodies have been first used more than one decade ago to visualize telomeric G-quadruplexes in ciliates.^[153] In 2013, an engineered antibody able to bind specifically and with high affinity to G-quadruplex structures in the human cells (BG4) was generated by Biffi *et al.* Immunostaining shows the presence of G-quadruplex structures in human genome and positional analysis revealed the formation of G-quadruplexes at telomeres and outside the telomeres. In particular, G-quadruplex structures are localized mostly outside the telomeres. However, immunodetection relies on a multistep protocol which requires the use of fixed and permeabilized cells, whose integrity is not perfect. Rodriguez *et al.* tried to circumvent this problem by targeting the G-quadruplex structures in the genome of living cells prior to cell fixing with a derivative of the ligand pyridostatin (see Chapter 6). In this way, the ligand binds to genomic G-quadruplexes in functioning cells. The detection requires cell fixation anyway and *in situ* click-chemistry in order to conjugate the PDS derivative with a fluorophore. The genomic sites of action of pyridostatin, which has been previously shown to produce replication- and transcription-dependent DNA damage, were mapped. Pyridostatin targeted mostly non-telomeric genomic sites and showed a propensity to interact with sequences with G-quadruplex forming potential.^[154] Another elegant method to identify G-quadruplex structures in living cells without fixation, permeabilization or chemical manipulation relies on targeting the G-quadruplex with quadruplex-selective fluorescent ligands acting as light-up probes. This strategy has been prominently applied in Chang labs and Monchaud labs.^[155,156] For further examples of G-quadruplex detection in living cells *via* smart light-up probe see also Chapters 6.2 and 6.3. The regions of the genome rich in G-quadruplex forming motifs will be described in the following.

3.1 G-quadruplex structures at telomeres

Telomeres are DNA-protein complexes located at the end of linear chromosomes in eukaryotic organisms.^[157] Telomeres protect the ends of chromosomes from homologous recombination and from being recognized as double strand breaks or damaged DNA and processed by the DNA repair machinery.^[158] Furthermore they are the substrate of the telomerase enzyme, a reverse transcriptase able to elongate them, thus providing a mechanism to solve the so-called “end-replication problem”.

Telomeric DNA

The telomeric DNA is formed by the repetition of a G-rich unit and comprises in humans a 10-15 kb long double-stranded region and a 150-250-nt long single-stranded 3'-overhang (**Figure 19a**).

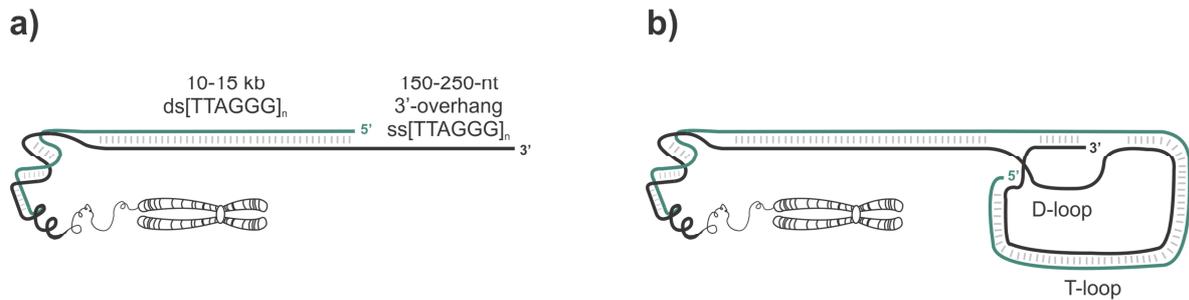


Figure 19 a) Structure of the human telomeres. b) T-loop and D-loop geometry.

This feature is conserved in the eukaryotic kingdom. The sequence of the G-rich repeat varies from organism to organism (see **Figure 20**) and is d(TTAGGG) in vertebrates. The “vertebrate motif” can be considered the ancestral telomeric motif in the animal kingdom and it is also found in the Choanozoa, a phylum of protists.

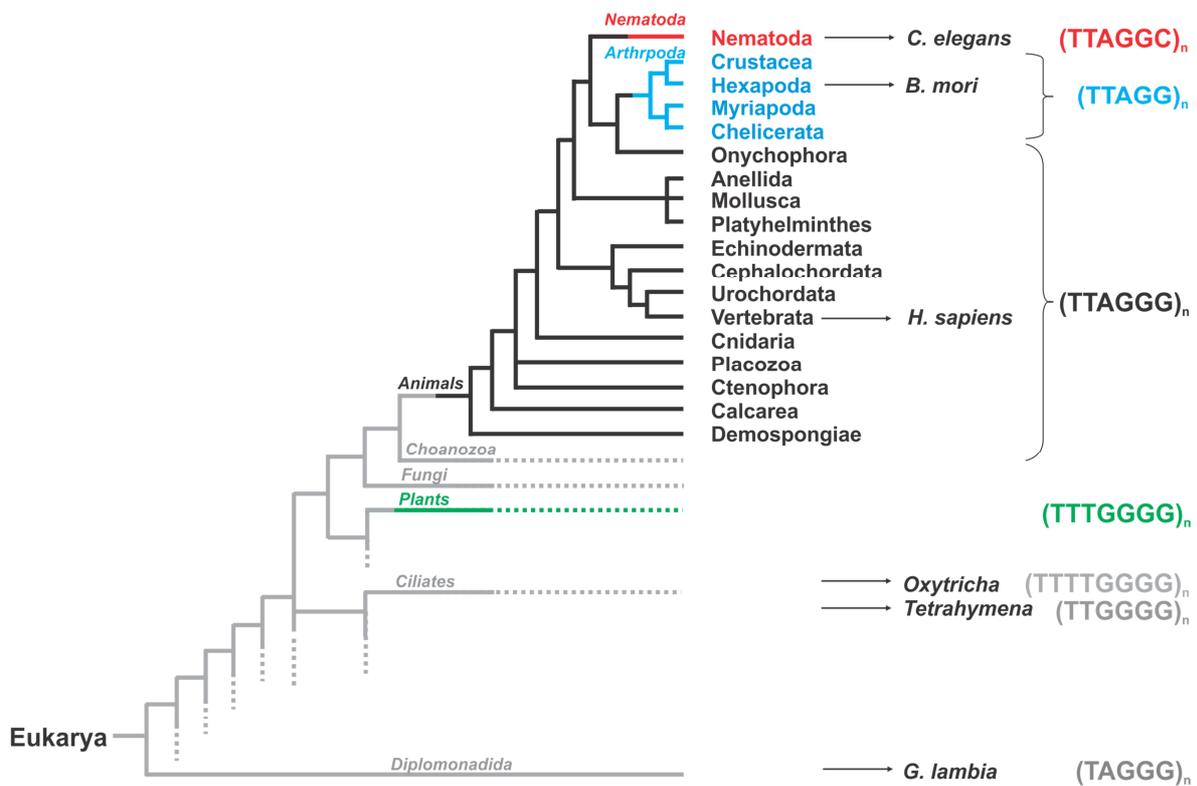


Figure 20 Distribution of telomeric motifs in the eukarya. Sequences of telomeric repeats derived from the most studied organisms are indicated on the right. The phylogenetic tree was adapted from references [159,160].

The presence of a canonical telomeric repeat is not always connected to significant telomerase activity or to the presence of telomerase at all. In fact, the telomerase enzyme has been lost, although rarely, in the evolution of plants (the onion, *Allium cepa*) and animals (the silkworm, *Bombyx mori*). Other species, for example the insect *Drosophila melanogaster*, do not possess any canonical telomeric motif.^[161]

The sequence of the 3'-end of human chromosomes is not precisely determined, while the sequence of the 5'-end is nearly always ATC-5'.

The single-stranded G-rich 3'-overhang has essentially two important functions: it recruits specific proteins (the so-called shelterin or telosome complex) to protect telomeres and it serves as primer for telomerase-mediated extension of telomeres. It can form alternative structures like G-quadruplexes (discussed in details in Chapter 2) or T-loops. The so-called T-loop (**Figure 19b**) is formed when the single-stranded portion folds back and invades the double-stranded telomeric DNA. The single-stranded DNA region is thus hidden from the DNA repair machinery, therefore this DNA architecture has been proposed to be a primordial mechanism for chromosome-end protection. As a consequence of the double-strand invasion, a portion of DNA remains single-stranded forming the so-called D-loop (**Figure 19b**). Little is known, however, about the dynamic behaviour of the T-loop and it is not clear whether it is persisting throughout the cell cycle.^[162,163]

Telomeres interacting proteins

Human telomeric DNA is packaged by histones into nucleosomes and bound to the telosome or shelterin complex (blue color code, **Figure 21a**), a sequence-specific complex of proteins that caps the chromosomes preventing illicit end-to-end joining and guarantees genomic stability.

The human shelterin complex is composed of six proteins: telomeric repeat-binding factor 1 (TRF1), telomeric repeat-binding factor 2 (TRF2), repressor and activator protein 1 (RAP1), TRF1-interacting nuclear protein 2 (TIN2), protection of telomeres 1 (POT1) and TPP1.

All together these proteins constitute the platform to recruit all the interaction partners involved in the telomeres maintenance and protection. The telomeric repeat binding factors TRF1 and TRF2 bind sequence-specifically as homodimers the double-stranded region of the telomeric DNA and play a role in the regulation of telomere length^[164] and in the stability of the T-loop structure.^[165]

TRF2 is supposed to be involved in tumorigenesis, since it is overexpressed in many human tumors.^[166] Critical loss of TTAGGG repeats or loss of TRF2 function results in the same cascade

of signals that is triggered by double-stranded DNA breaks. Mammalian RAP1 is not able to bind telomeric DNA and localizes at the telomers via binding to TRF2. In mice, TRF1- and TPP1-deficiency leads to 100% perinatal mortality, while lack of RAP1 protein does not affect mice viability, even though it increases telomeres fragility.^[167] TIN2 tethers TRF1, TRF2 and TPP1.

Moreover, it has a major role in stabilizing the TPP1/POT1 complex on the single-stranded DNA.^[168] TPP1 and POT1 are localized at the single-stranded overhang *via* the specific binding of

POT1 to the single-stranded decamer TTAGGGTTAG. POT1 comprises two Oligonucleotide/Oligosaccharide Binding (OB) fold domains: OB1 recognizes 5'-TTAGGG while OB2 binds to the downstream TTAG-3'. POT1 is supposed to prevent the binding of replication protein A (RPA) to the single stranded DNA. RPA is a ssDNA binding protein essential for triggering DNA damage response. It has been shown that G-quadruplex formation enhances the POT1/TPP1 ability to block the access of RPA to the telomeres.^[169] TPP1 increases by 10-fold the DNA affinity of POT1 and recruits the telomerase enzyme to the tip of the chromosomes. TPP1

and POT1 enhance together the processivity of the telomerase.^[170] The POT1-TPP1 *in vitro* complex is extremely dynamic and slides back and forth on the telomeres overhang, possibly inducing folding and refolding of G-quadruplex structures. This sliding mechanism may result in an enhancement of the telomerase processivity.^[171] Extra-telomeric function of shelterin proteins have been also reported.^[157]

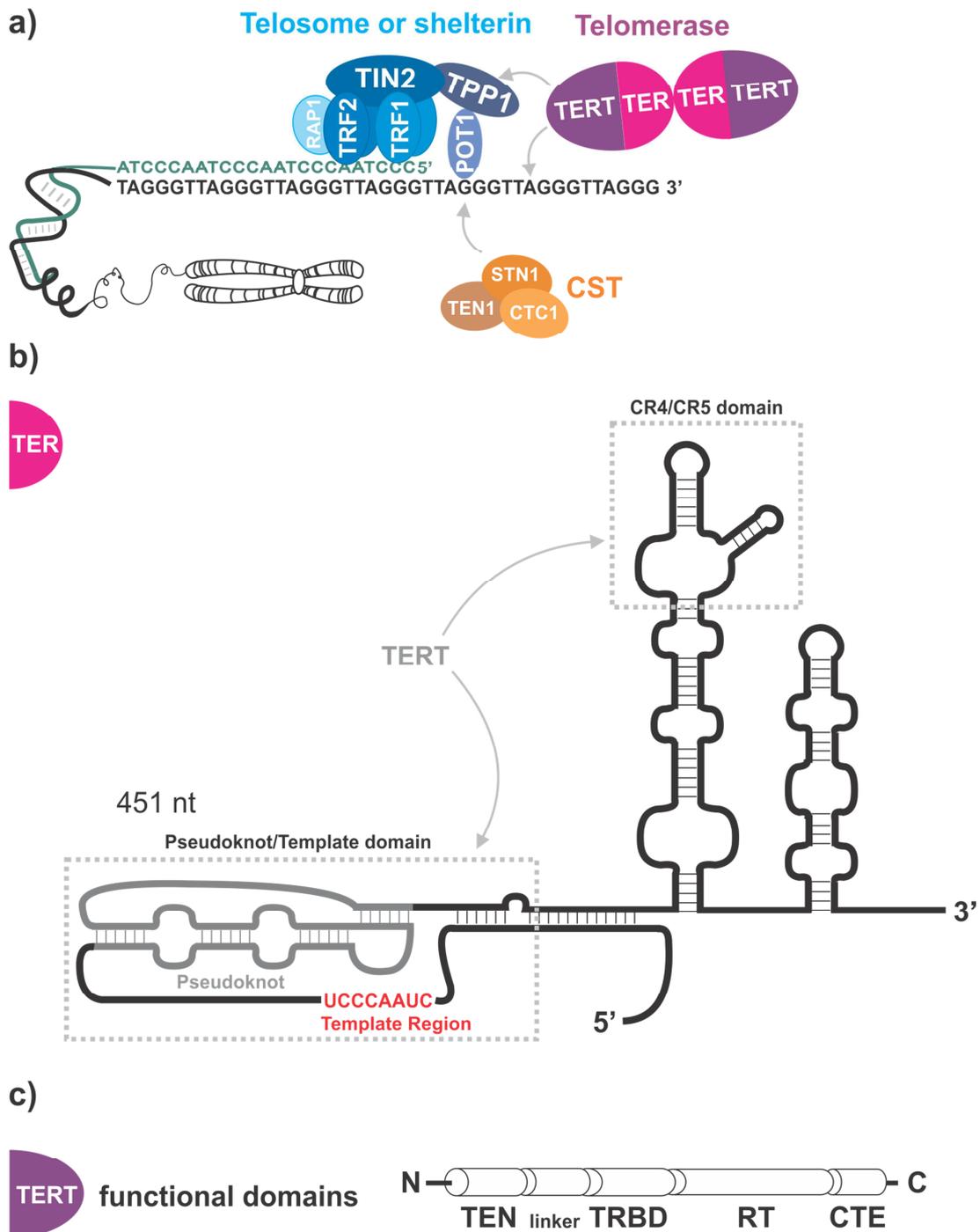


Figure 21 a) Human telomeres interacting proteins. b) Scheme of the secondary structure of the telomerase TER subunit. c) Functional domains of the TERT subunit. Adapted from reference [172].

CST

The human CST (CTC1-STN1-TEN1) complex (orange color code, **Figure 21a**) is composed of the three proteins CTC1 (conserved telomere maintenance component 1), STN1 (suppressor of Cdc thirteen) and TEN1 (protein involved in Telomeric pathways in association with STN1, number 1). The CST is able to bind the single stranded telomeric DNA via multiple OB-folding domains. It has been proposed that, at least in yeasts, the structure of CST resembles the structure of the replication protein A (RPA) complex.^[173] The CST complex is supposed to be implicated in supporting the protection of the telomeres and to inhibit telomerase activity by competing with POT1/TPP1 complex and sequestration of the telomerase primer.^[174]

Telomerase

The telomerase enzyme (purple color code, **Figure 21a**) in its minimal form is composed of a protein subunit (the telomerase reverse transcriptase TERT), and an RNA component (the telomerase RNA TER), containing the template used by TERT to add multiple DNA repeats to the 3'-overhang of telomeres. The structures available for TERT and TER derived from different organisms are reviewed by Sandin and Rhodes in reference [172].

Size and sequence of each component vary from species to species. For example, the size of the RNA subunit TER ranges from 150 nt in ciliates to 1300 nt in yeasts, while the length of TER in vertebrates is about 450 nt (**Figure 21b**). The two structural domains framed in **Figure 21b** with gray boxes are conserved: a stem-loop element (CR4-CR5) and the catalytically essential pseudoknot-template domain. These two domains are directly involved in the interaction with the TERT subunit (gray arrows in **Figure 21b**).^[175]

The TERT subunit is conserved among many organisms and is composed of about 1100 amino acids (**Figure 21c**). The essential N-terminal domain (TEN) is binding with high affinity the single stranded telomeric DNA and is recruited by TPP1/TIN2 in the shelterin complex. It is connected via a flexible linker to the RNA bonding domain (TRBD), which is able to interact with the CR4/CR5 element of TER. The central reverse transcriptase domain (RT) is flanked by the TRBD domain and a C-terminal domain (CTE). The 21 Å resolution 3D structure obtained by single particle EM (electron microscopy) suggests that human telomerase enzyme exists as a dimer composed of two TER subunits (each 153 kDa) connected by a flexible hinge and two TERT subunits (each 127 kDa) at the periphery, locating the catalytic pockets about 180 Å far from each other.^[176] EM structural analysis reveals the presence of an "open-state" and a "closed-state". Sauerwald *et al.* speculate also that the telomerase exists as a dimer to extend in parallel two different telomeric substrates from sister chromatids.

Telomere maintenance

50 years ago, in the early days of the telomere research, the "end-replication problem" was already clear: according to the semi-conservative DNA replication mechanism, each round of cell division causes the shortening of telomeres because conventional DNA polymerases are not able to replicate the 3'-ends of linear chromosomes. The central role of telomeres to copy completely the end of chromosomes was proposed at the beginning of the Seventies; however the details for the replication mechanism were still unclear. Telomerase was identified the first time in 1985 in the

ciliate *Tetrahymena* by Greider and Blackburn.^[177] A few years later Szostak discovered a gene coding for a protein essential telomere maintenance in *Saccharomyces cerevisiae*.^[178] The pioneering work by Greider, Blackburn and Szostak was awarded in 2009 with the Nobel Prize in Physiology or Medicine “for the discovery of how chromosomes are protected by telomeres and the enzyme telomerase”.

The telomerase enzyme circumvents the “end-replication problem” by adding in a processive way DNA repeats to the 3'-overhang of telomeres.

A proposed mechanism for telomere elongation is schematically represented in **Figure 22**. In humans the enzyme telomerase is recruited to the telomeres by the protein TPP1. The TERT subunit adopts a closed-ring configuration and does hardly undergo any conformational change during the telomere synthesis. Once the enzyme is loaded on the telomeres, the RNA template partially pairs with the single stranded telomeric DNA, facilitating the alignment of the 3'-end in the active site. After catalysing the addition of one telomeric repeat (5'-GGTTAG-3' in humans), the telomeric overhang transiently dissociates from the RNA template region and allows base pairing of the newly synthesized DNA overhang with the distal region of the RNA template. A new telomeric repeat is then added. The translocation/elongation cycle is reiterated in a processive way. The TEN domain shows affinity for single-stranded telomeric DNA and is supposed to bind upstream of the telomeric 3'-end contributing to anchor the telomerase to the chromosome and potentially facilitating the telomerase processivity by maintaining the association during the translocation step.^[179] However, the exact mechanism of re-alignment of DNA overhang with the RNA template after one round of elongation is unknown.

The telomerase processivity is the capability of adding multiple repeats without releasing the DNA substrate. Under homeostatic conditions, during the S phase of the cell cycle, one molecule of telomerase adds processively 60 nt. When telomeres are elongating (non-equilibrium conditions) multiple molecules of telomerase are involved and the extension is less processive.^[180] Recently published results based on single-molecule data suggest that the telomerase rate of extension is 1 nt/sec (60 nt/min) and is 6 times faster in presence of POT1/TPP1.^[181]

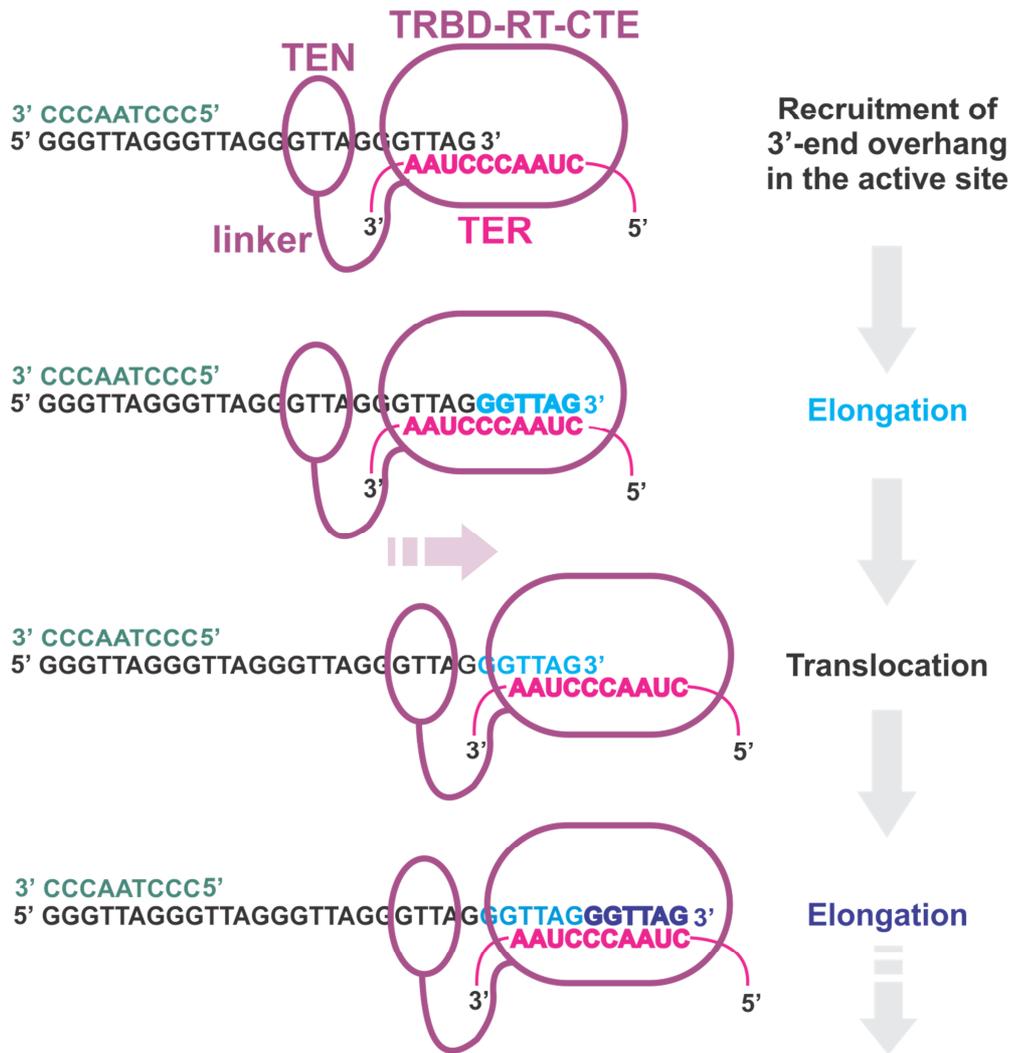


Figure 22 Mechanism of telomeres elongation catalysed by telomerase enzyme. Adapted from reference [160].

In normal human cells, after successive cellular divisions the telomeres progressively shorten and after reaching a critical length, cellular senescence is provoked. Therefore the progressive shortening of telomeres is supposed to be a “molecular clock”.^[182] On the other side, cancer cells proliferate indefinitely and are considered immortal. Telomeres play a central role in determining the replicative potential of cells; in particular immortal cells maintain telomeres at a stable length. The level of telomerase expression changes through life: it is maximum in the early stage of embryonic development and in pluripotent stem cells and it is drastically reduced, although detectable, in adult somatic cells, where telomeres are shortening with aging (**Figure 23**).

Telomerase enzyme is re-activated in 85-90% of all human tumors.^[183] Tumors have a longer proliferative history, therefore cancer cells are usually featured by short telomeres, whose maintenance above a critical minimal length is guaranteed by high levels of telomerase activity.

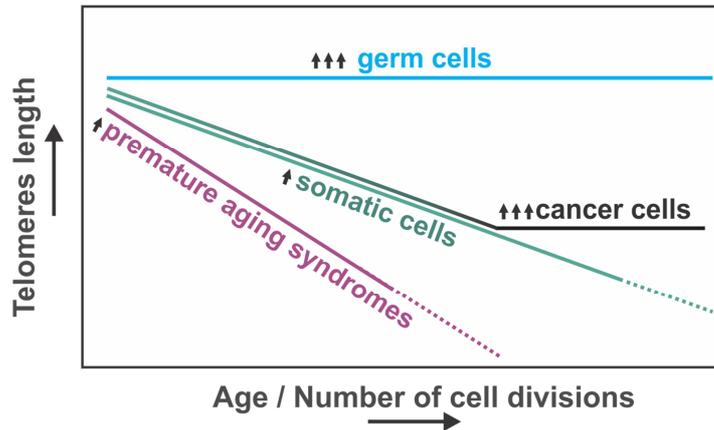


Figure 23 Changes in telomere length of different classes of cells over time. Number of arrows is proportional to the level of telomerase activity. The full line represents viable cells and the dotted line indicates apoptosis/chromosomal instability/arrest of the cell cycle.

Although most of the human tumors are connected to high levels of telomerase activity, and rely on a telomerase-based mechanism for telomeres elongation, about 5% of human tumors exploit a different mechanism to overcome replicative mortality.^[184] The so-called alternative lengthening of telomeres (ALT) pathway is thought to be based on homologous recombination. Evidence of the presence of 5' C-rich telomeric overhang in the nematode *Caenorhabditis elegans* suggested the necessity to reconsider the architecture of mammalian telomeres and indeed 5' C-rich telomeric overhangs were found to be prevalent in tumor cells relying on the ALT pathway.^[185] Although *C. elegans* has been often chosen as a model organism to investigate ALT mechanism, NMR structural investigations have recently shown that, at least *in vitro*, the 5'- and 3'-overhang architecture of *C. elegans* telomeres is completely different from other multicellular eukaryotes, underlying possible limitations in the use of this organism as model.^[186]

3.2 G-quadruplex structures in transcription

One or more potential G-quadruplex forming sequences have been identified within 1000 nt upstream of the transcription start site (TSS) of 50% of human genes. In particular, promoters of human oncogenes and regulatory elements are more likely to contain G-quadruplex forming sequences than housekeeping or tumor suppressor genes.^[148]

Figure 24 summarizes possible mechanisms of transcriptional regulation *via* G-quadruplex (GQ) elements in gene promoters. The formation of G-quadruplex on the template strand might cause a stalling of the replication machine, while G-quadruplex formation on the coding strand could in principle contribute to maintain the transcribed strand in a single stranded conformation and enhance the transcription. Furthermore, the formation of G-quadruplex structures on the coding strand could also recruit transcriptional repressor or transcriptional enhancers. See also the recent review by Rhodes and Lipps for more details on the regulatory roles of G-quadruplex structures.^[187]

a) Transcription

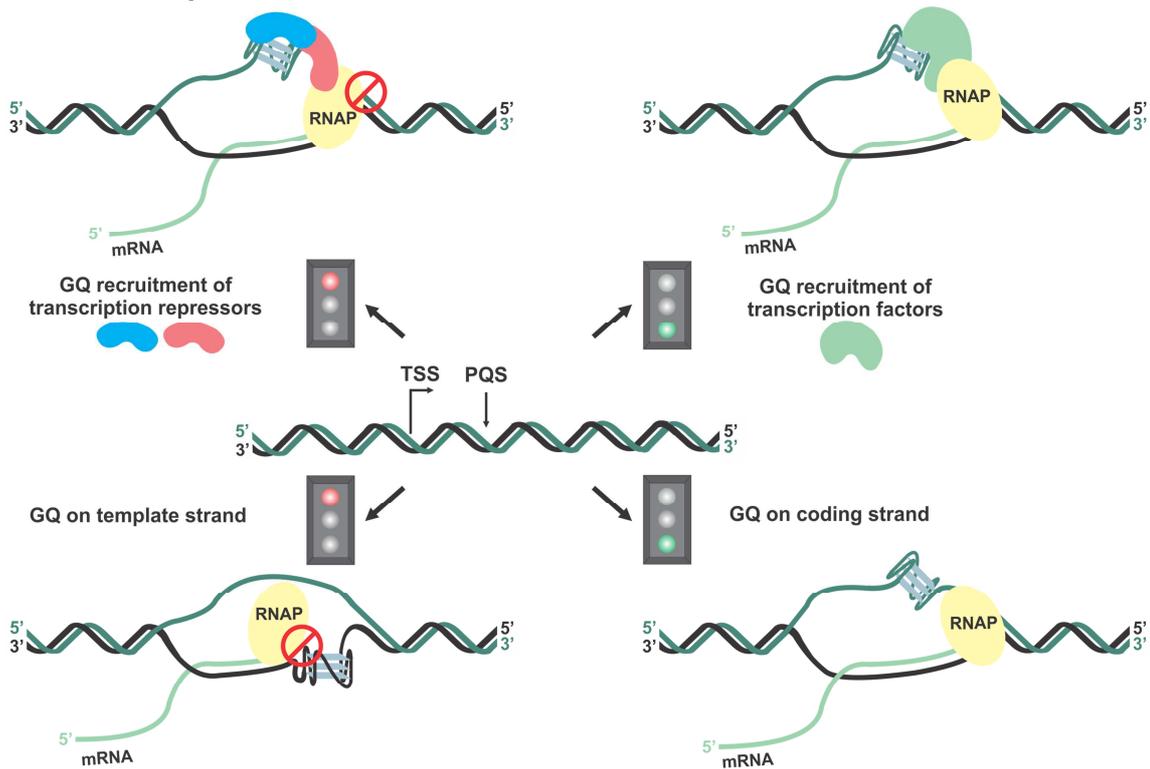


Figure 24 Putative biological roles of the G-quadruplex (GQ) in transcription, as proposed by Bochman *et al.*^[188] depicted as cartoon scheme. RNA polymerase (RNAP) is indicated in yellow. The transcription start site (TSS) and the location of the potential G-quadruplex forming sequence (PQS) are indicated with arrows.

Potential G-quadruplex forming sequences (PQS) have been identified in the promoter regions of the genes encoding for the following proteins: MYC, KIT receptor tyrosine kinase, B cell lymphoma 2 (BCL-2), vascular endothelial growth factor (VEGF), hypoxia-inducible factor 1 α (HIF1 α), the transcription factor c-MYC, platelet-derived growth factor α polypeptide (PDGF α), PDGF receptor β polypeptide (PDGFR β), KRAS, retinoblastoma protein 1 (RB1), human telomerase reverse transcriptase (h-TERT).

c-MYC

The transcription factor MYC is involved in transcriptional regulation pathways related to a broad spectrum of cellular activities, such as cell growth, metabolism and survival.^[189] MYC is in fact able to bind to 10-15% of all promoter regions.^[190] c-MYC is the most frequently amplified oncogene in human cancers and its expression is deregulated by gene amplifications in solid tumors and by chromosomal translocation and leukemias and lymphomas.^[191] Transcription of c-MYC is primarily regulated by a 27 nt G-rich sequence present within the nuclease hypersensitivity element III₁ (NHE III₁).^[192] This sequence, located -142 to -115 bp upstream of the P1 promoter of c-MYC oncogene, exists in an equilibrium between transcriptionally active forms (double helical and single stranded) and a silenced form able to adopt G-quadruplex structures.^[193-195] The role of G-quadruplexes in the regulation of c-MYC transcription will be discussed in detail in Chapter 4.1.

The purine-rich 27 nt-long sequence responsible of *c-MYC* regulation, indicated as *mycPu27* (**Table 4**), contains five G-runs (I-V). Various G-quadruplex structures derived from different G-rich tracts of the *c-MYC* regulatory element NHE III1 have been reported until now.^[95,196–198]

It has been shown that in a K⁺-containing solution only the four consecutive 3' G-runs (II-V) are involved in the formation of the major G-quadruplex structure which is causing the gene transcriptional silencing.^[199,200] The 22 nt-long wild-type sequence containing the four 3' G-runs (*myc2345*, **Table 4**) adopts a parallel stranded, propeller topology and forms a mixture of four loop isomers which are in dynamic equilibrium.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
<i>mycPu27</i>	T	G	G	G	G	A	G	G	G	T	G	G	G	G	A	G	G	G	T	G	G	G	G	A	A	G	G
			I				II				III				IV				V								
<i>myc2345</i>				T	G	A	G	G	G	T	G	G	G	G	A	G	G	G	T	G	G	G	G	A	A		
<i>mycPu22_G14T/G23T</i>				T	G	A	G	G	G	T	G	G	G	T	A	G	G	G	T	G	G	G	T	A	A		
<i>myc_Pu22_G11T/G23T</i>				T	G	A	G	G	G	T	T	G	G	G	A	G	G	G	T	G	G	G	T	A	A		
<i>mycPu22_G14T/G20T</i>				T	G	A	G	G	G	T	G	G	G	T	A	G	G	G	T	T	G	G	G	A	A		
<i>mycPu22_G11T/G20T</i>				T	G	A	G	G	G	T	T	G	G	T	A	G	G	G	T	T	G	G	G	A	A		
<i>mycPu24_G13I</i>				T	G	A	G	G	G	T	G	G	I	G	A	G	G	G	T	G	G	G	G	A	A	G	G
<i>myc1234</i>	T	G	G	G	G	A	G	G	G	T	G	G	G	G	A	G	G	G	T								
<i>mycPu19_G2A/G11A</i>	T	A	G	G	G	A	G	G	G	T	A	G	G	G	A	G	G	G	T								
<i>mycPu19_G2A/G14A</i>	T	A	G	G	G	A	G	G	G	T	G	G	G	A	A	G	G	G	T								
<i>myc1245</i>	T	G	G	G	G	A	G	G	G	T	T	T	T	A	G	G	G	T	G	G	G	G	A				

Table 4 Overview of the various sequences derived from the NHE III₁ of *c-MYC*

The biological function of the *c-MYC* promoter quadruplex depends on the equilibrium between the four loop isomers and each loop isomer contributes to its inhibitory effect. The structures of the different loop isomers can be obtained by dual G-to-T substitutions (*mycPu22_G14T/G23T*, *mycPu22_G11T/G23T*, *mycPu22_G14T/G20T*, *mycPu22_G11T/G20T*, **Table 4**) and differ in the arrangement of one- and two- nucleotide loops.^[200] The major conformer is the one formed by the sequence *mycPu22_G14T/G23T* (**Table 4**) and its solution NMR structure has been solved in 2005 by Yang and co-workers.^[198] This structure has been used in the G-quadruplex ligand interaction studies presented in Chapters 6.2 and 6.3. As reported by Yang,^[198] the TAA-3' and 5'-TGA flanking sequences are able to form multiple stable conformations. Thus, this sequence is present in solution as a mixture of a major conformation and one or more minor conformations which are likely featured by the same parallel G-quadruplex core, but different capping structures. The exact structure of these minor conformations has not been characterized until now. The pioneering work of Siddiqui-Jain *et al.* showed that small molecules stabilizing the *c-MYC* G-quadruplex can reduce *c-MYC* transcription in cancer cells.^[201–203] Since then, several classes of small molecules able to bind and stabilize *c-MYC* quadruplex have been developed^[199,203–215] and few of them have been structurally characterized in complex with *c-MYC* G-quadruplex by NMR.^[197,216,217]

Alternative structures have been proposed by Phan *et al.* (*mycPu24_G13I* PDB: 2A5P and *myc2345* – no PDB). The major G-quadruplex formed under superhelicity conditions involves the

four consecutive 5' G-runs (I-IV).^[34] The wild-type sequence containing the four 5' G-runs (*myc1234*, **Table 4**) adopts a mixture of two different loop isomers which can be isolated by dual G-to-A substitutions (*mycPu19_G2A/G11A* and *mycPu19_G2A/G14A*, **Table 4**). The conformation of the main loop isomer (*mycPu19_G2A/G11A*) is featured again by an all-parallel core with propeller loops.

c-KIT

The *c-KIT* proto-oncogene encodes for a membrane-bound glycoprotein belonging to the family of tyrosine kinase receptors involved in developmental and growth processes. Gain-of-function and point mutations are the most common *c-KIT* mutations detected in human malignancies like human gastrointestinal stromal tumor (GIST), mastocytosis, acute myeloid leukaemia, sinonasal NK/T-cell lymphoma, germ cell tumors and melanoma.^[218] Two G-rich sequences have been identified in the promoter region of the human *c-KIT* gene, so-called *c-kit1* d(AGGGAGGGCGCTGGGAGGAGGG), occurring between -87 and -109 base pairs relative to the TSS, and *c-kit2* d(CGGGCGGGCGCGAGGGAGGGG), located between -140 and -160 base pairs relative to the TSS. X-ray and crystal structures are available for *c-kit1*,^[219,220] while several models have been proposed on the basis of NMR data for the folding of *c-kit2*.^[221,222] The *c-kit1* crystal structure (PDB: 3QXR) was solved in 2012 and it was the first X-ray structure published for a gene promoter G-quadruplex. In contrast to the relative abundance of X-ray structures of telomeric G-quadruplexes, very few crystal structures have been reported until now for gene promoter quadruplexes. Interestingly, X-ray and NMR structures of *c-kit1* are in good agreement, while the folding topologies of the human telomeric quadruplex are strongly dependent on the experimental method and on the experimental conditions employed.

BCL-2

BCL-2 is a membrane protein which plays an important role in the maintenance of mitochondrial integrity and it has an anti-apoptotic effect. The *BCL-2* oncogene is overexpressed in a number of human tumors, such as B-cell and T-cell lymphomas, breast cancer and gastric carcinoma.^[223] The 5'-end of the P1 promoter of the human *BCL-2* oncogene includes a 39-nt GC-rich element containing 6 G-tracts of 3-5 consecutive guanines and can potentially form 15 different G-quadruplexes. The structure of the G-quadruplex formed by the mid G-tracts, which appears to be the most stable in K⁺-containing solution, was determined in 2006 by NMR^[224] and an alternative structure involving the G-tracts number 1, 2, 4 and 5, featured by a 13-nt loop has been recently proposed.^[92] The C-rich counterpart has been extensively studied and is able to form i-motif structures.^[225-227]

VEGF

The vascular endothelial growth factor (VEGF) is a pluripotent cytokine and a potent angiogenic factor, consisting of two identical subunits binding to VEGF receptors on the surface of endothelial cells.^[228,229] VEGF is up-regulated in many tumors and several VEGF inhibitors against tumor-mediated angiogenesis are in clinical trials.^[230] The region located between -85 and -50 bp relative to the transcription start site is essential for *VEGF* promoter activity in many human cancer cell lines and it has been reported to form G-quadruplexes.^[231] The solution NMR structure of a 22

nt sequence derived from the *VEGF* promoter sequence by G to T mutation of two residues in the central loop was recently reported.^[232]

HIF1 α

The hypoxia-inducible factor 1 (HIF-1) is a transcription factor which regulates the response to the deprivation of oxygen. It is composed of two subunits: the α subunit is inducible by hypoxia (HIF1 α) and the β subunit is constitutively expressed (HIF1 β). HIF-1 is up-regulated in several tumor types and promotes the formation of an aggressive tumor phenotype.^[233] The *HIF1 α* promoter region contains a polyPu:polyPy rich sequence able to form G-quadruplex structures and is located between -85 and -65 base pairs upstream of the TSS and overlapping putative binding sites for activating transcription factors.^[234]

c-MYB

The *c-MYB* proto-oncogene encodes for a transcription factor involved in the regulation of differentiation, survival and proliferation of cells in the hematopoietic system and in the gastrointestinal tract. Deregulation of *c-MYB* expression seems to activate genes related to cancer progression and metastasis.^[235] The *c-MYB* promoter contains three regions of four GGA repeats located 17 base pairs downstream of the TSS. The structure of the G-quadruplex formed by d[(GGA)₄] is featured by two G-tetrads and contains an unusual G(:A):G(:A):G(:A):G heptad.^[94] It was solved by NMR long before the identification of the d[(GGA)₄] motif in the *c-MYB* promoter.^[236]

PDGF α

Platelet-derived growth factors have been implicated in the pathogenesis of a number of tumor types, via autocrine stimulation of cancer cells, angiogenesis and control of interstitial fluid pressure.^[237]

The 5'-flanking region of the gene encoding for the platelet-derived growth factor α polypeptide (PDGF α) contains one nuclease hypersensitive element (NHE) located between -120 and -33 base pairs relative to the TSS and is characterized by high GC-content. The G-rich strand of the NHE of *PDGF α* contains five G-tracts, each separated from the neighbouring by single nucleotide. It has been proposed that this region is able to form two distinct stable G-quadruplex structures, each containing four G-tetrads and characterized by an all-parallel strands arrangement.^[238]

PDGFR β

The 38 nt-long G-rich strand of the NHE in the promoter region of the gene encoding for the platelet-derived growth factor receptor β polypeptide (PDGFR β) contains seven G-tracts of 2-7 consecutive guanines. DMS footprinting and mutational studies allowed the identification of the residues involved in the formation of the major intramolecular G-quadruplex formed in the *PDGFR β* promoter region and NMR studies highlighted the presence of a broken G-tract with a snapback motif.^[239]

KRAS

KRAS encodes for a GTPase protein acting as a signal transducer (RAS) playing a key role in regulation of cell growth. *KRAS* is the sixth most mutated gene in all cancers.^[240] The NHE in the promoter region of *KRAS* contains a G-rich strand that has been proposed to assume two different conformations.^[241]

RB1

The retinoblastoma gene *RB1* encodes for a nuclear phosphoprotein that acts as tumor suppressor and is functionally inactivated in most human cancers.^[242] Formation of a G-quadruplex at the 5'-end of the *RB1* gene has been shown by Sugiyama *et al.*^[243]

TERT

The catalytic subunit of the human telomerase reverse transcriptase (h-TERT) elongates the 3'-overhang of human telomeres as described in Chapter 3.1 (see **Figure 22** and **Figure 23**). Phan and co-workers identified a region with high propensity for G-quadruplex formation between -20 and -110 bp relative to the TSS. This region includes 14 G-tracts able to form multiple G-quadruplex structures and the fragment located between -41 and -60 bp relative to the TSS resulted suitable for NMR structural characterization, which revealed the coexistence of two major conformations in equilibrium.^[244]

The majority of the G-quadruplex structures reported for promoter oncogenes were derived from mutant sequences (mutations are highlighted in red in **Table 5**) optimized in order to stabilize a major conformation.

In fact, many G-rich sequences extracted from gene promoters contain more than 4 G-tracts composed of more than 3 consecutive guanines, therefore more G-quadruplex structures can potentially be formed and coexist in equilibrium with each other. This confers to the system a dynamic character, which could be determinant for the biological function, as already reported for *c-MYC*.^[199] *PDGF α* , *PDGF β* and *BCL-2* promoters contain also multiple overlapping G-quadruplex forming G-tracts.

Parallel-stranded structures with three tetrads are the most common folding topologies observed in G-rich sequences derived from the promoter sequences of human oncogenes (**Table 5**). They all feature a 1:n:1 loop arrangement, with $2 \leq n \leq 5$ and propeller/double chain reversal being the most favourite loop geometry. However, structures with only two tetrads (*c-MYB*, PDB: 1MYQ) or four tetrads (*PDGF α* , model proposed on the basis of CD data and dimethyl sulfate footprinting) were also reported. There are also several examples of promoter G-quadruplex structures featured by different folding topologies, containing loop in geometries other than propeller and snapback motifs. Interestingly, the structure of *c-kit1*, reported both by NMR and X-ray crystallography (PDB: 2O3M, 4WO2), presents a unique arrangement embedding an isolated non-G-tract guanine into a G-quartet. Moreover, a 3'-end snapback guanine has been reported to be involved in the formation of an external tetrad in the NMR structures determined for *c-MYC* by Phan *et al.* (PDB: 2A5P) and in the NMR-based model proposed for the G-quadruplex derived from the *PDGF β* promoter region.^[239] The (3+1) strand arrangement has been observed only in the promoters derived from *h-TERT* (PDB: 2KZD) and *BCL-2* (PDB: 2KF8U).

Table 5 summarizes the structures available up to date for G-quadruplexes formed in promoter regions of oncogene.

Gene	Sequence	Ligand	PDB code	Method	Topology
BCL-2	d(GGGCGCGGGAGGAAT <u>TTGGGCGGG</u>)	-	2F8U	NMR	(3+1), (p,l-long, l)
B-raf	d(GGGCGGGGAGGGGAAGGG <u>A</u>)	-	4H29	X-ray	interlocked dimer, 7 quartets
c-MYC	d(TGAGGGTGGI <u>GAGGGTGGGGAAGG</u>)	-	2A5P	NMR	snapback motif (p, p, p, d-long)
	d(TGAGGGTGGG <u>TAGGGTGGGTAA</u>)	-	1XAV	NMR	all parallel, (p, p, p)
	d(TAGGGAGGGT <u>AGGGAGGGT</u>)	-	2LBY	NMR	all parallel, (p, p, p)
	d(TGAGGGTGGT <u>GAGGGTGGGGAAGG</u>)	Phen-DC3	2MGN	NMR	see 2A5P
	d(TGAGGGTGGG <u>TAGGGTGGGTAA</u>)	quindoline	2L7V	NMR	see 1XAV
	d(GAGGGTGGI <u>GAGGGTGGGGAAGG</u>)	TmPyP4	2A5R	NMR	see 2A5P
h-TERT	d(A <u>GGG</u> I A <u>GGGG</u> CT <u>GGG</u> A <u>GGG</u> C)	-	2KZD	NMR	(3+1), (l, l, p)
	d(A <u>IGGG</u> <u>AGGG</u> I CT <u>GGG</u> A <u>GGG</u> C)	-	2KZE	NMR	All parallel, (p, p, p)
c-KIT (c-kit1)	d(AGGGAGGGCGC ^{Br} <u>UGGGAGGAGGG</u>)	-	3QXR	X-ray	dimer, all parallel, snapback motif (p, p, l, l-long)
	d(AGGGAGGGCGCT <u>GGGAGGAGGG</u>)	-	2O3M	NMR	all parallel, snapback motif (p, p, l, l-long)
	d(AGGGAGGGCGCT <u>GGGAGGAGGG</u>)	-	4WO2	X-ray	as 3QXR
c-KIT (c-kit2)	d(CGGGCGGGC <u>ACGAGGGAGGGT</u>)	-	2KQG	NMR	all parallel, (p, p-long, p)
	d(CGGGCGGGCGCG <u>AGGGAGGGT</u>)	-	2KQH	NMR	all parallel (p, p-long, p)
	d(CGGGCGGGCGCG <u>AGGGAGGGT</u>)	-	2KYO	NMR	interlocked dimer, all parallel
	d(CGGGCGGGCGC <u>TAGGGAGGGT</u>)	-	2KYP	NMR	all parallel, (p, p-long, p)
RET	d(GGGGCGGGGCGGGGCGGGG <u>T</u>)	-	2L88	NMR	all parallel, (p, p, p)
c-MYB	d(GGAGGAGGAGGA)	-	1MYQ	NMR	dimer, 2-tetrad
VEGF	d(CGGGGCGGGCC <u>TTGGGCGGGT</u>)	-	2M27	NMR	all parallel, (p, p-long, p)

Table 5 Survey of the G-quadruplex structures from promoter regions of oncogenes, deposited in the Protein Data Bank up to January 2016. Mutations compared to the wild type sequence are highlighted in red. The guanine residues involved in the formation of the G-quartets are underlined in the sequence. The topology is described on the basis of the relative orientations of the strands (all parallel, 2+2 or 3+1) and the geometry of the loops encountered in direction 5' to 3' (l, d and p indicate lateral, diagonal and propeller loop, respectively).

3.3 G-quadruplex structures in replication

Immunostaining of human cells showed a maximum number of foci during the S phase, consistent with replication-dependent formation of G-quadruplex structures.^[152] Besnard *et al.* showed that G-quadruplex-forming motifs are associated with the majority of replication origins and they speculated that G-quadruplex structures might favour duplex unwinding in negatively supercoiled DNA and facilitate the loading of origin recognition complex (ORC) and other replication factors.^[245]

During DNA replication, replicative helicases separate the two DNA strands and, while the leading strand is synthesized continuously, the synthesis of the lagging strand is discontinuous and lets the lagging strand transiently single-stranded. G-quadruplex structures can be formed, in particular on the lagging strand, stalling the replication machine (**Figure 25**).

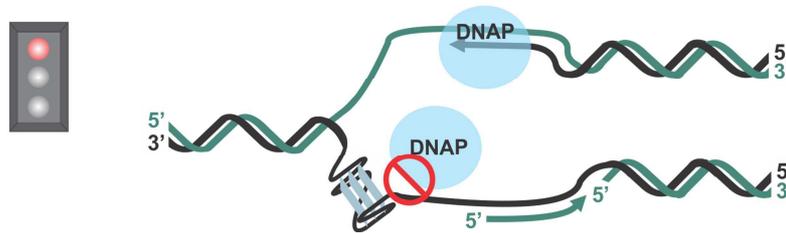


Figure 25 Putative biological roles of the G-quadruplex structures in replication, as proposed by Bochman *et al.*^[188] depicted as cartoon scheme. DNA polymerase (DNAP) is indicated in light blue.

G-quadruplex structures must be resolved in order to allow the polymerase to use the DNA strand as template. In fact, many DNA helicases have been shown to bind and/or unwind G-quadruplex structures *in vitro*, including human helicases like WRN, BLM, FANCD1 and PIF1, associated with genomic instability and increased cancer risk.^[188] In cell lines from human patient affected by Fanconi anemia, the loss of G-quadruplex unwinding capability is correlated with genomic deletions overlapping regions with high G-quadruplex formation potential.^[246] This study provides an important piece of evidence connecting human diseases to the loss of capability to unwind G-quadruplex. Furthermore Paeschke *et al.* suggested that the resolution of G-quadruplex structures by PIF1 helicase suppresses genome instability.^[247]

3.4 G-quadruplexes at other locations

Minisatellites are tandem repeats of 10-100 nt motifs, often very unstable in human germlines. The subtelomeric minisatellite CEB1 is destabilized in meiosis^[248] and the solution structure of a G-rich fragment derived from CEB1 has been characterized by NMR (PDB: 2MB4).^[91] G-quadruplex forming sequences have also been identified downstream of the TSS, in the first intron and on the non-template strand.^[249] NMR structures of the G-rich sequences derived from the 5'-intron of the human *CHL1* gene and from the first intron the *N-MYC* gene have been reported.^[90,250]

Recent studies showed that there are a number of putative G-quadruplexes forming sequences in the mitochondrial DNA (mtDNA) and indeed G-quadruplex were detected in living cells using an engineered fluorescent carbazole ligand able to bind G-quadruplexes and accumulate selectively in mitochondria.^[155]

4 G-quadruplex as novel anti-cancer target

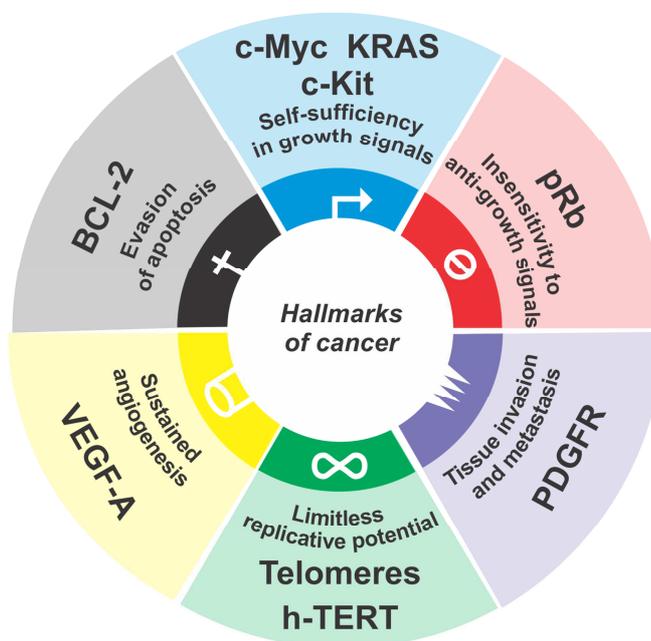


Figure 26 The six functional capabilities which are proposed to be hallmarks of cancer according to Weinberg and Hanahan. G-quadruplexes forming oncogenes and chromosomal regions associated to each of the six processes are listed. Adapted from References [194,251,252].

Weinberg and Hanahan proposed that most of cancers are featured by six essential alterations in cell physiology and acquire during their development the same set of functional capabilities: self-sufficiency in growth signals, insensitivity to growth-inhibitory signals, evasion of apoptosis, tissue invasion and metastasis, limitless replicative potential and sustained angiogenesis.^[251] Interestingly, G-quadruplexes have been identified in genes and chromosomal regions that are represented in each of the six cancer hallmarks (**Figure 26**).^[252]

There is general interest in developing drugs able to target molecular processes that are more active in cancer than in normal cells.^[253]

G-quadruplexes are structurally diverse and have unique molecular recognition features that make G-quadruplex binders suitable drugs with potentially high specificity for cancer cells. In fact, structural studies (see Chapter 2.3) revealed that the G-quadruplex architecture is highly heterogeneous due to diversity of loop geometries and strands orientation and, from a drug-development point of view, its globular fold makes it closer to a protein molecular receptor than the traditional target, the B-DNA form.

The B-DNA was historically the first target for many of the drugs that have been used in anti-cancer therapy. Alkylating agents, in particular inter-strand cross-linking agents like cisplatin^[254] have been extensively used and have been proved to be potent and efficacious in treating patients affected by cancer. However, the clinical potential of these DNA-damaging agents

is weakened by their general cytotoxicity and non-specificity. In contrast to alkylating agents, anti-tumor antibiotics (e.g., doxorubicin,^[255] a non-specific DNA intercalator that causes topoisomerase II inhibition) seem to have a modest selectivity for cancer cells over normal cells, most likely due to interaction with protein-DNA complexes. Code-reading drugs, able to selectively bind in a sequence-specific manner the major groove (e.g., triplex-forming oligonucleotides -TFOs-^[256] and peptide nucleic acids -PNAs^[257]) or the minor groove (e.g., netropsin and distamycin)^[258] via donor-acceptor interactions and/or shape complementarity, have also been developed.

The discovery of G-quadruplex structures, with their polymorphisms, provided a novel class of molecular targets for DNA-interactive compounds and offered new opportunities for drug design. Up to date there is only one G-quadruplex ligand that has successfully entered the phase II of Clinical trials (Quarfloxin, CX-3543, see Chapter 6, **Figure 42c**).^[207] However, the development of G-quadruplex ligands is considered a promising area in cancer therapeutics, especially as part of a rational synergistic anti-cancer strategy.^[259] “Synthetic lethality” refers to the condition where, given two genes, mutation or chemical inhibition of either one gene alone is compatible with viability, while mutation of both genes is lethal.^[260] This concept was exploited to develop cancer-specific cytotoxic agents that in combination generated an enhanced inhibitory effect. For example, synergistic interaction of telomestatin (see Chapter 6, **Figure 42p**) and WRN helicase inhibitor caused 70% inhibition of cancer cell growth.^[261] Moreover, the pyridostatin (see Chapter 6, **Figure 42o**) antiproliferative effect exhibited synergy with a DNA-dependent protein kinase (DNA-PKcs) inhibitor.^[262] An important factor for an effective synergistic action seems to be the order of administration of the drugs.

Furthermore, G-quadruplex ligands may be employed to target at the same time oncogenic and telomeric G-quadruplexes (dual targeting), which could result in additional therapeutic advantage.

4.1 Targeting DNA G-quadruplexes with small molecules

The first therapeutic application of G-quadruplex binding ligands was the inhibition of the telomerase enzyme. This strategy was based on the observation that the telomerase enzyme is up-regulated in 85-90 % of all cancers, while its expression level is drastically reduced in healthy cells. Because of this level of specificity, targeting telomerase and telomeres maintenance is nowadays considered an attractive platform for anticancer drug discovery.^[263–265] Some aspects regarding targeting G-quadruplexes as telomere-directed anticancer drug modality will be discussed in the following. An overview of small molecule G-quadruplex binders will be given in more detail in Chapter 6.

The initial idea underlying the G-quadruplex-mediated telomerase inhibition was to stabilize or to induce formation of G-quadruplex structures at the single stranded 3'-overhang of telomeres in order to prevent the access of telomerase to its substrate/primer. Indeed, already in 1991 it was shown that the telomerase activity is inhibited *in vitro* if the substrate contains pre-folded G-quadruplex structures.^[266] The first G-quadruplex ligand able to inhibit effectively the telomerase enzyme was reported in 1997 by Neidle and Hurley laboratories.^[267]

Initially the pharmaceutical industry and the academic world were sceptical about this novel strategy because of the time-lag expected for telomerase inhibition, assuming that telomeres need to be eroded down to a critical low point before any phenotypic effect occurs. It was argued that in contrast to conventional chemotherapeutic agents, small molecules that target G-quadruplexes in order to inhibit telomerase may not induce cytotoxicity immediately after administration. Knowing that about 100 nt per round of replication are lost, cells with a mean telomere length of 5 kb and a 24 h cell doubling time would reach the critical telomere length at which the cell enters the senescence in 40-50 days.^[263] However, subsequent studies showed that anti-tumor effects can occur much faster (after 10 days) and without telomere shortening.^[268,269]

It is now accepted that a G-quadruplex ligand binding to telomeric DNA competes with hPOT1 and facilitates the uncapping of telomerase and hPOT1, initiating the DNA damage response mechanisms, which are fast and lead to senescence and apoptosis (**Figure 27**).^[270] It has to be mentioned that recently published results from Balci and co-workers suggest that *in vitro* the formation of G-quadruplex at the 3'-single stranded telomeric overhang might contribute to favour the loading of the POT1-TPP1 proteins and suppress the binding of the replication protein A (RPA protein) which recognizes the single stranded DNA as a damaged site.^[169]

G-quadruplex binding ligands cannot be considered "simple" telomerase inhibitors, but telomere targeting agents acting with more than one mechanism and able to induce short-term and long-term responses.^[271] As suggested in a recent review by Ohnmacht and Neidle, the expression "telomere targeting" should be used to indicate the behaviour of well-studied G-quadruplex binders like RHPS4^[269] and BRACO-19^[268] (see Chapter 6) that have been shown to bind and stabilize single-stranded telomeric overhang and to induce hPOT1 and hTERT uncapping, leading to a fast DNA damage response.^[264]

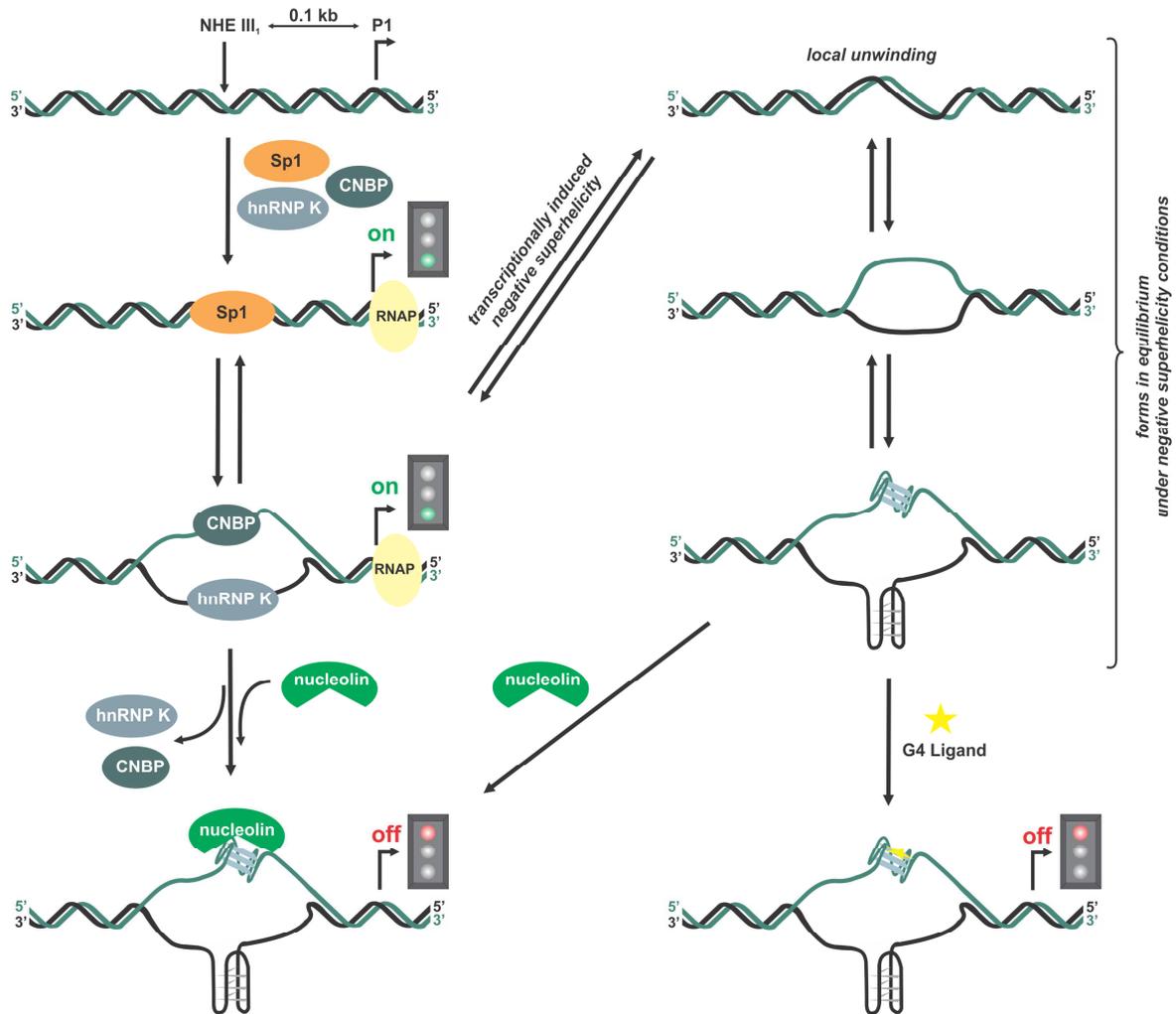


Figure 28 Cartoon scheme of the proposed model for activation/silencing of *c-MYC*.^[204] Partially adapted from references [252,276].

The role of G-quadruplex in gene regulation has been extensively studied in the *c-MYC* system.^[276] **Figure 28** summarizes the proposed mechanisms of *c-MYC* regulation via G-quadruplex structures that can be formed in the NHE III₁ of the promoter region.

As already discussed, the negative superhelicity plays an important role in formation of G-quadruplex structures and under negative superhelicity the *c-MYC* NHE III₁, situated around 130 nt upstream of the promoter region P1, has been shown to co-exist as duplex, partially melted and non-B-DNA secondary conformation (**Figure 28**).

The NHE III₁ element is the binding substrate for several proteins. The transcription factor Sp1 (orange code, **Figure 28**) can bind GC-rich double helical regions but it is not binding the *c-MYC* promoter in quiescent cells. Sp1 binds the NHE III₁ element after induction of *c-MYC* transcription by serum stimulation and acts as trans-activator. hnRNP K (heterogeneous nuclear ribonuclear protein K, light gray code, **Figure 28**) and CNBP (cellular nucleic acid binding protein, dark green code, **Figure 28**) have been shown to bind respectively to the C-rich and G-rich single-stranded region of the NHE III₁ element and activate the transcription.^[277,278] Nucleolin is a multifunctional

phosphoprotein that has been shown to selectively bind *in vitro* with high affinity and selectivity to the *c-MYC* G-quadruplex structure (green code, **Figure 28**). Overexpression of nucleolin results in a reduction of the activity of the *c-MYC* promoter.^[279] When the *c-MYC* NHE III₁ element is in its G-quadruplex/i-motif form, the substrate of Sp1, CNBP and hnRNP K is not accessible anymore and the transcription is silenced. In this model the G-quadruplex acts as a silencer element. Stabilization of the G-quadruplex form by nucleolin results in negative regulation of *c-MYC* transcription.^[279] A landmark paper by Hurley and co-workers showed that point mutations destabilizing the G-quadruplex resulted in an increased transcriptional activity, while G-quadruplex stabilizing compounds such as TMPyP4 (yellow star, **Figure 28**) down-regulate the *c-MYC* expression.^[280]

Notably, the compound Quarfloxin, initially developed to inhibit *c-MYC* expression (see also Chapters 3.2 and 6), acts as inhibitor of the RNA polymerase I in the transcription of the human ribosomal RNA gene (*rRNA*), which is over-expressed in cancer cells. *rRNA* is GC-rich and several PQS regions have been identified in the coding strand. G-quadruplex structures in the coding strand are supposed to prevent the renaturation of the DNA and increase the transcription rate of RNA polymerase I. Moreover, the nucleolar protein nucleolin, which is essential for the ribosomal RNA (*rRNA*) synthesis, has nanomolar affinity for G-quadruplex structures and may contribute to increase the polymerase transcription rate stabilizing the *rRNA* gene G-quadruplexes. Drygin *et al.* showed that Quarfloxin concentrates in the nucleolus and is able to disrupt the nucleolin/*rRNA* G-quadruplex complex. This results in the selective inhibition of RNA polymerase I and suppression of the ribosome biogenesis, which is deregulated in cancer cells. Furthermore, the nucleolin is redistributed from the nucleolus to the nucleoplasm, mimicking stress conditions and leading to apoptosis.^[207]

4.2 Occurrence and regulation by RNA G-quadruplexes

RNA G-quadruplexes and their biological role will be shortly discussed in this chapter (for further details, see review [281]).

Early reports on formation of RNA G-quadruplexes date back to the 1990s,^[282] as well as the first studies establishing a connection between RNA G-quadruplexes and biological processes. Explicit evidence for the formation of RNA G-quadruplexes in cell has been recently provided in the Balasubramanian lab using G-quadruplex structure-specific antibodies.^[283]

Bioinformatics studies have shown that 5'- and 3'-untranslated regions of mRNAs are enriched in G-quadruplex-forming sequences, suggesting that these elements could be involved in RNA processing or translational control.^[249,284] Furthermore, it has been recently shown that telomeres are not transcriptionally silent but they are transcribed into TERRA RNAs by the DNA-dependent RNA polymerase II and are supposed to be involved in chromatin remodelling and regulation of telomerase activity.^[285-287] TERRA transcripts are composed of tandem repeats of r(UUAGGG)

with heterogeneous length (100 to 900-nt) and can fold into G-quadruplex structures (PDB: 2M18, 2KBP, 3IBK, 3MIJ).^[288–291]

Interestingly, while DNA G-quadruplexes seem to adopt a diversity of folding topologies, the currently available structures indicate that RNA G-quadruplexes prefer an all-parallel folding with propeller loops, irrespective of the flanking nucleotides and the experimental conditions.^[292]

Biophysical studies proved that RNA G-quadruplexes are thermodynamically more stable than DNA G-quadruplexes^[293] and the structural rationale for this stability has been proposed by Neidle and co-workers on the basis of X-ray data and molecular modelling.^[290] The 2'-hydroxyl group interacts with phosphate as well as backbone oxygen atoms and with polar group from the bases. These interactions affect the hydration structure, reducing the number of structural water molecules and strengthening the hydrogen bonding network of the constrained water molecules. These extensive intramolecular contacts confer stability to the overall structure. Furthermore, on the basis of these additional hydrogen bonding options, targeting selectively RNA vs DNA G-quadruplex may be achieved with appropriate design of the ligand.^[294]

The role of 5'-UTR RNA G-quadruplex structures in translation regulation has been reviewed by Bugaut and Balasubramanian^[281] and the two proposed mechanisms are summarized in **Figure 29**.

It has been suggested that, in the context of the cap-dependent translation, the formation of G-quadruplex structures at the 5'-UTR can interfere with the formation of the pre-initiation ribosomal complex or interrupt the scanning mechanism (**Figure 29a**). In fact, G-quadruplex elements have been shown to act as translation repressors in several mRNAs.^[295–297] An alternative mechanism of translation initiation (cap-independent translation) uses internal ribosomal entry sites (IRES) and by-passes the 5'-cap. The IRES-mediated translation has been identified in several cellular mRNAs, such as the 5'-UTR of the human vascular endothelial growth factor (*hVEGF*). Basu and co-workers have recently reported that a G-quadruplex-forming element in the IRES domain of the 5'-UTR mRNA of *hVEGF* directly recruits the 40S ribosomal subunit and is essential for IRES-mediated translation (**Figure 29b**).^[298,299] Therefore, RNA G-quadruplexes have no “universal 'translational suppressor' effect”.^[300]

According to the regulation mechanisms proposed above, targeting RNA G-quadruplexes at the 5'-UTR with small molecules may allow to modulate the translation and, ultimately, to control gene expression.

However, in contrast to the number of studies reported until now on the DNA G-quadruplex interaction with ligands (see Chapter 6), there are only few reports on small molecules targeting RNA G-quadruplex and their mechanism of action is still elusive.^[281]

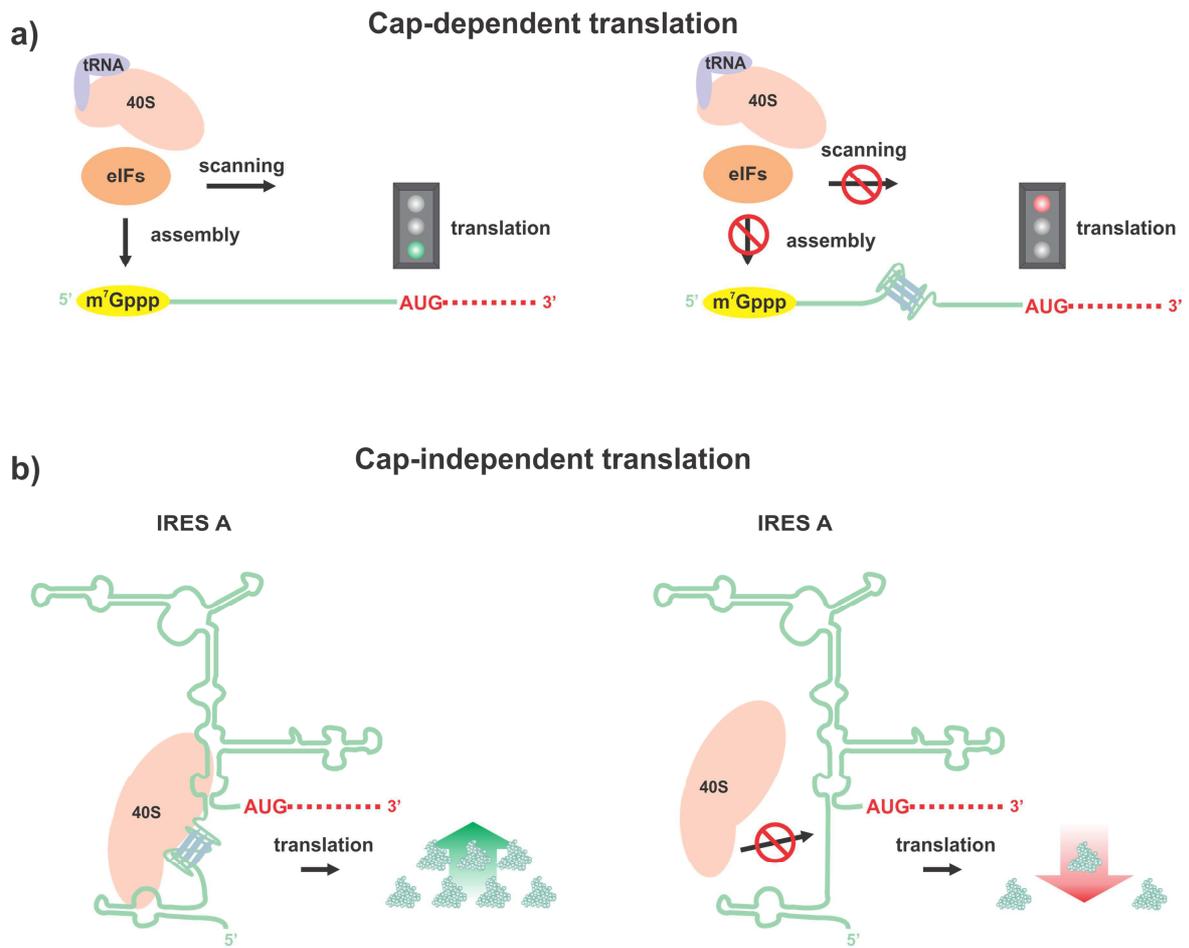


Figure 29 Putative role of RNA G-quadruplex structures at 5'-UTR in a) cap-dependent and b) cap-independent translation. Adapted from references a) [281] and b) [298].

5 NMR of G-quadruplexes

Atomic-resolution structures of G-quadruplexes have been determined in the last two decades by means of X-ray crystallography and solution state NMR spectroscopy (**Figure 30a**). The increase in number of DNA G-quadruplex structures determined by NMR has been quite large, compared to the growth of the overall DNA structures determined by NMR (**Figure 30b**).

NMR provides information not only about the structure at atomic resolution in solution, but also allows to investigate the dynamics and the kinetics of the G-quadruplex and its interaction with proteins and ligands. NMR methods to investigate G-quadruplex structures have been reviewed by Webba da Silva^[301] and, more recently, by Heddi and Phan.^[302]

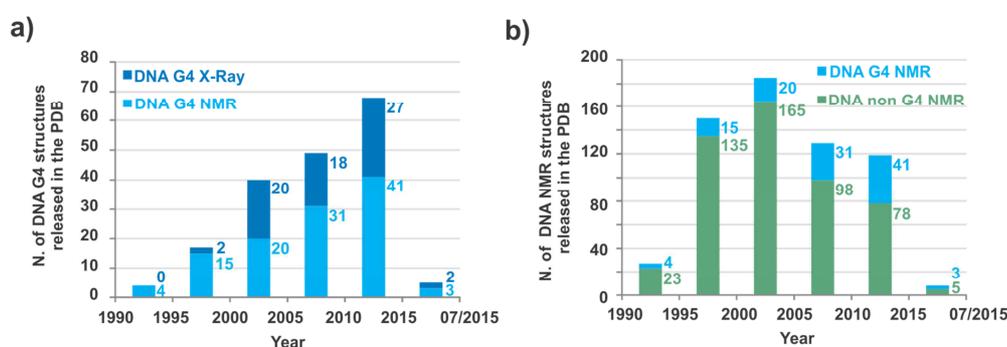


Figure 30 Number of DNA structures released in the PDB since 1990 a) containing only DNA in G-quadruplex conformation (G4, query: "quadruplex") and determined by NMR (light blue) or by X-ray crystallography (blue), b) containing only DNA in G-quadruplex conformation (G4, query: "quadruplex", light blue), or only DNA in other conformations (green).

5.1 Favouring a single G-quadruplex conformation

As discussed in Chapter 2.3, G-quadruplex polymorphism in solution is high and the presence of multiple conformations is revealed in a 1D ^1H NMR spectrum by detecting a number of imino proton peaks exceeding the number of guanine residues involved in the formation of a single quadruplex. This complicates the analysis of the NMR spectra and the determination of high resolution NMR structure of G-quadruplexes. Hence, different strategies have been developed to favour a specific conformation and ultimately improve the quality of the spectra. The folding topology of a G-quadruplex can be controlled first of all by adjusting the nature and the concentration of the stabilizing cation and/or by varying the molecular crowding conditions.^[114,125,137,303] Moreover, mutations at the 3'- and/or 5'-flanking sequences have been shown to shift the conformational equilibrium towards a single conformer. For example, the human telomeric sequence d[TTGGG(TTAGGG)₃A] (Tel24) used in the kinetic studies (see Chapter 7) presents an A→T mutation at position 2 and a T→A mutation at position 24 compared to the human telomeric wild-type sequence. These two mutations guarantee the formation of a major

conformer (95% populated) suitable for NMR determination.^[304] Mutations in the loop regions can also influence the relative populations of more conformers in equilibrium, as reported, e.g. for the structural determination of the major G-quadruplex formed in the sequences derived from the NHE III₁ element in the promoter region of *c-MYC* (see Chapter 3.2).^[198] Rational incorporation of chemically modified nucleotides (**Figure 31**) has been also extensively used to control the conformational equilibria of G-quadruplex in solution and decrease the structural polymorphism.^[305]

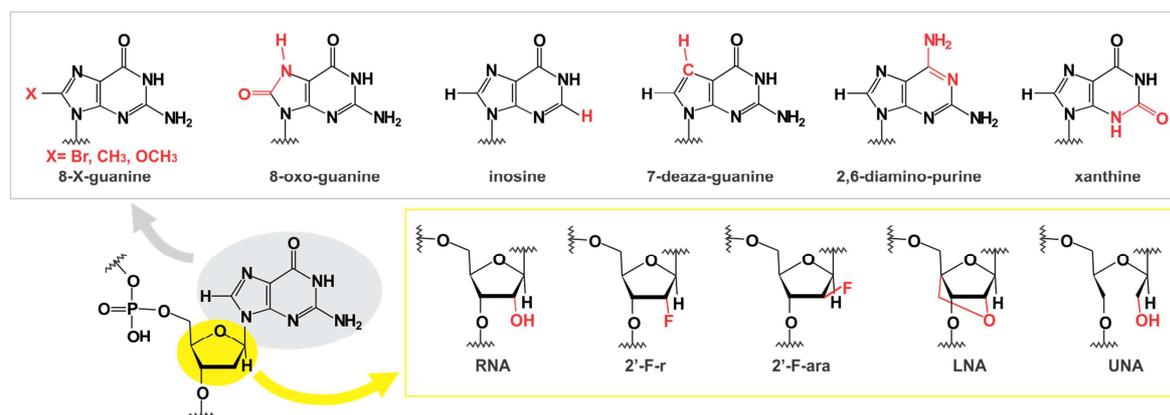


Figure 31 2'-deoxyguanosine modifications affecting the guanine moiety (gray box) or the deoxyribose moiety (yellow box).

In general, bulky substituents at position 8 tend to stabilize *syn* guanosine conformation. The influence of C8-substitutions on the Tel24 G-quadruplex overall stability has been systematically investigated by Lech *et al.*^[306] 8-Br-substitutions on *syn* guanines result always in an increased thermal stability compared to the unmodified sequence. Furthermore, a bromine atom facing a wide groove has a stronger stabilizing effect than a bromine atom in a medium groove. In the latter case, the presence of a propeller loop bridging the groove contributes to reduce the stabilizing effect. Karsisiotis *et al.* have explained these effects in terms of steric hindrance and proposed that the general effect of an 8-Br-substitution is to prevent the formation of narrow grooves.^[305] The effect of 8-oxo-substitution is more drastic, since it results in a non-planar 5-membered ring which impairs the G-quadruplex stability and possibly modifies its folding topology. Hartig and co-workers suggested that a combination of xanthine (X) and 8-oxo-guanine (O) in a complementary manner may favour the formation of a X·O·X·O tetrad which could be incorporated in a quadruplex with little loss of stability.^[307] Phan and co-workers have recently shown how a rational incorporation of xanthine and 8-oxo-guanine can be used either to replace a G·G couple in a G-quartet without altering the overall folding^[308] or to invert the G-quartet polarity.^[84]

Replacement of guanine with its analogue inosine has been used not only to influence the conformational equilibria, but also to facilitate the NMR assignment. For the same purposes, replacement of thymine with 6-Br-uracile has been employed in several NMR studies.

Modifications can be introduced also in the sugar moiety. 2'-deoxy-2'-fluoro-arabinoguanosine (2'-F-araG) has been shown to favour *anti* conformation of the glycosidic bond and replacement of *anti* 2'-deoxyguanosine with 2'-F-arabinoguanosine results in an increase of the G-quadruplex

stability.^[309–311] It has been shown that 2'-F-arabinose adopts a south-east sugar pucker (see **Figure 2**),^[312] therefore 2'-F-araG tends to adopt an *anti* conformation to avoid steric repulsion between the nucleobase and the β -fluorine. 2'-deoxy-2'-fluoro-guanosine (2'-F-r) modification is also well tolerated at *anti* positions, while it leads to perturbation of the folding topology when substituted at *syn* positions.^[313] Interestingly, a selective replacement of *syn* guanine with the *anti*-favoring 2'-deoxy-2'-fluoro-guanosine (2'-F-r) can result in the inversion of the Hoogsteen hydrogen bonding polarity of a specific tetrad, without alteration of the overall G-quadruplex folding.^[83] Locked nucleic acids (LNA) and unlocked nucleic acids (UNA) were also used as tool to manipulate G-quadruplex polymorphism. In LNA the O2' is linked to the C4' via a methylene bridge which locks the furanose ring in N-type conformation as usually observed in RNA. On the other hand, the sugar moiety in UNA residues is very flexible, due to the lack of the bond connecting C3' and C4'. LNA substitutions at *anti* positions are generally stabilizing, while substitutions at *syn* positions are disrupting the folding topology.^[310] A combination of LNA and UNA substitutions has been used by Plavec and co-workers to favour the formation of one predominant conformation in the G-rich vascular endothelial growth factor (VEGF) aptamer.^[314] NMR studies have shown that oligonucleotides composed only by LNA can form G-quadruplexes with all the sugarpuckers in C3'-endo^[89,315] and recently, the first crystal structure of an all-LNA G-quadruplex has been reported.^[316]

5.2 NMR parameters: chemical shifts and coupling constants

The imino protons belonging to guanine residues involved in a G-quartet *via* Hoogsteen base pairing typically resonate between 10.5 and 12.2 ppm and they are clearly separated from the signals stemming from canonical Watson-Crick base pairing, clustered between 12 and 15 ppm. Furthermore, the imino proton signals originating from C⁺-C symmetric base pair in the i-motif conformation are clustered around 15.5 ppm. A 1D ¹H spectrum in water allows already determining whether the investigated oligonucleotide is forming a G-quadruplex and how many tetrads and/or different conformations are formed. The jump-return-echo water suppression scheme with an excitation maximum at around 11.5 ppm allows to detect imino signals with optimal sensitivity.^[317] The number of imino proton signals detected between 10.5 and 12.2 ppm is indicative of the number of guanine residues involved in G-quadruplex formation and, consequently, of the number of tetrads composing the quadruplex. Additional canonical base pairs may be formed by the residues belonging to the loops and the flanking sequences constituting the capping structures. Further, the imino signals from these capping structures are usually shifted downfield compared to the imino signals from the G-tetrad moiety.

Figure 32 displays the imino region of the 1D ¹H spectrum recorded after addition of the C-rich (green) complementary strand to a G-rich telomeric sequence pre-folded in G-quadruplex structure (black). The hybridization process is slow and allows to detect simultaneously 1) the G-quadruplex unfolding, 2) the i-motif formed by the C-rich sequence at acidic pH and 3) the nascent duplex,

product of the hybridization of the two complementary strands. Interestingly, there is no overlap between the imino signals arising from the three different structures.

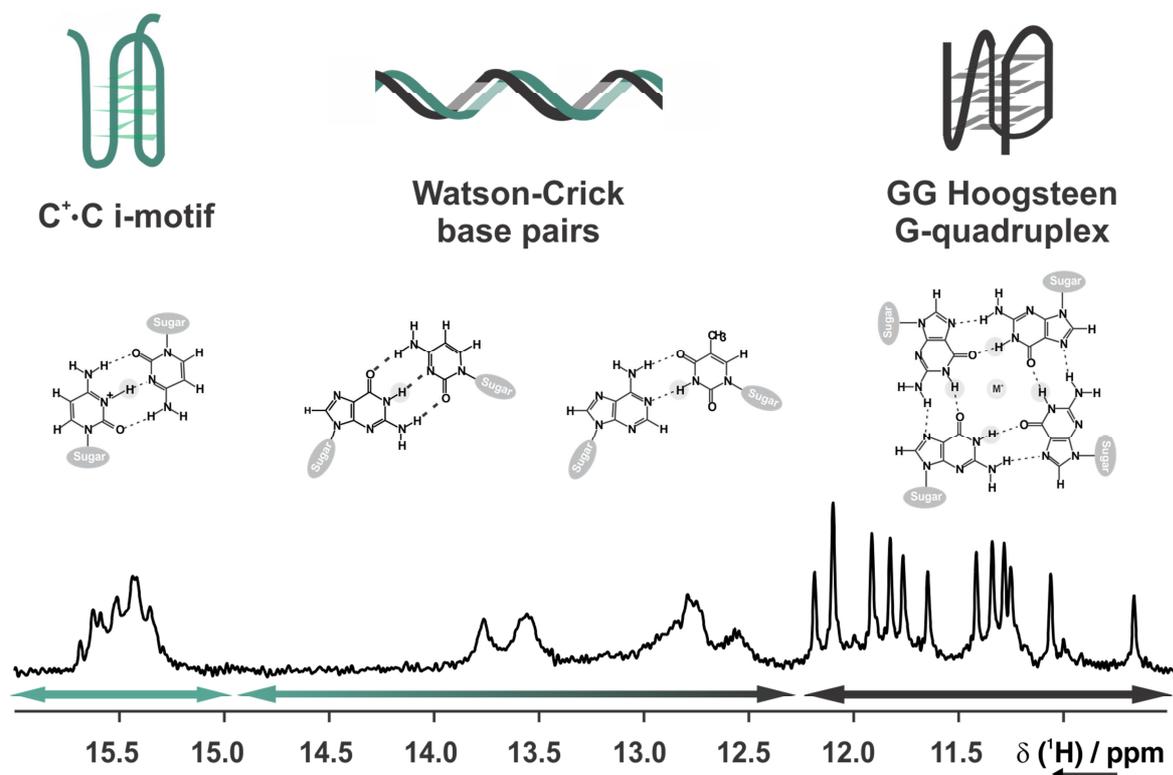


Figure 32 Imino region of 1D ^1H spectrum recorded to monitor the hybridization of telomeric G-rich (black) and C-rich (green) complementary strands.

Hydrogen to deuterium exchange will result in the fast exchange of the imino protons located on the external tetrads, while protons belonging to the inner tetrads are more protected from exchange with solvent and are usually detectable at room temperature for several hours. In the presence of capping structures, even imino protons from the external tetrads can be detected after H-D exchange.

The ^{13}C chemical shift gives also valuable information, allowing to distinguish in the aromatic region between H2 protons of adenine residues, resonating around 152 ppm, and the other aromatic protons (**Figure 33**). Greene *et al.* have systematically investigated the effect of the glycosidic torsion angle on the ^{13}C and ^{15}N chemical shifts of guanosine residues in bimolecular G-quadruplexes derived from telomeric sequences. ^{13}C chemical shifts of C8, C1', C2', C3', and C4' have been shown to be correlated to the glycosidic torsion angle χ .^[83,318] Assignment of the guanine anomeric carbon C1' and aromatic carbon C8 in the hybrid-1 G-quadruplex formed by the telomeric sequence Tel24 is indicated in **Figure 33a-b**. The C1' and C8 carbons belonging to guanine residues in *syn* conformation (violet) are on average downfield shifted compared to the one observed for *anti* guanine residues (gray).

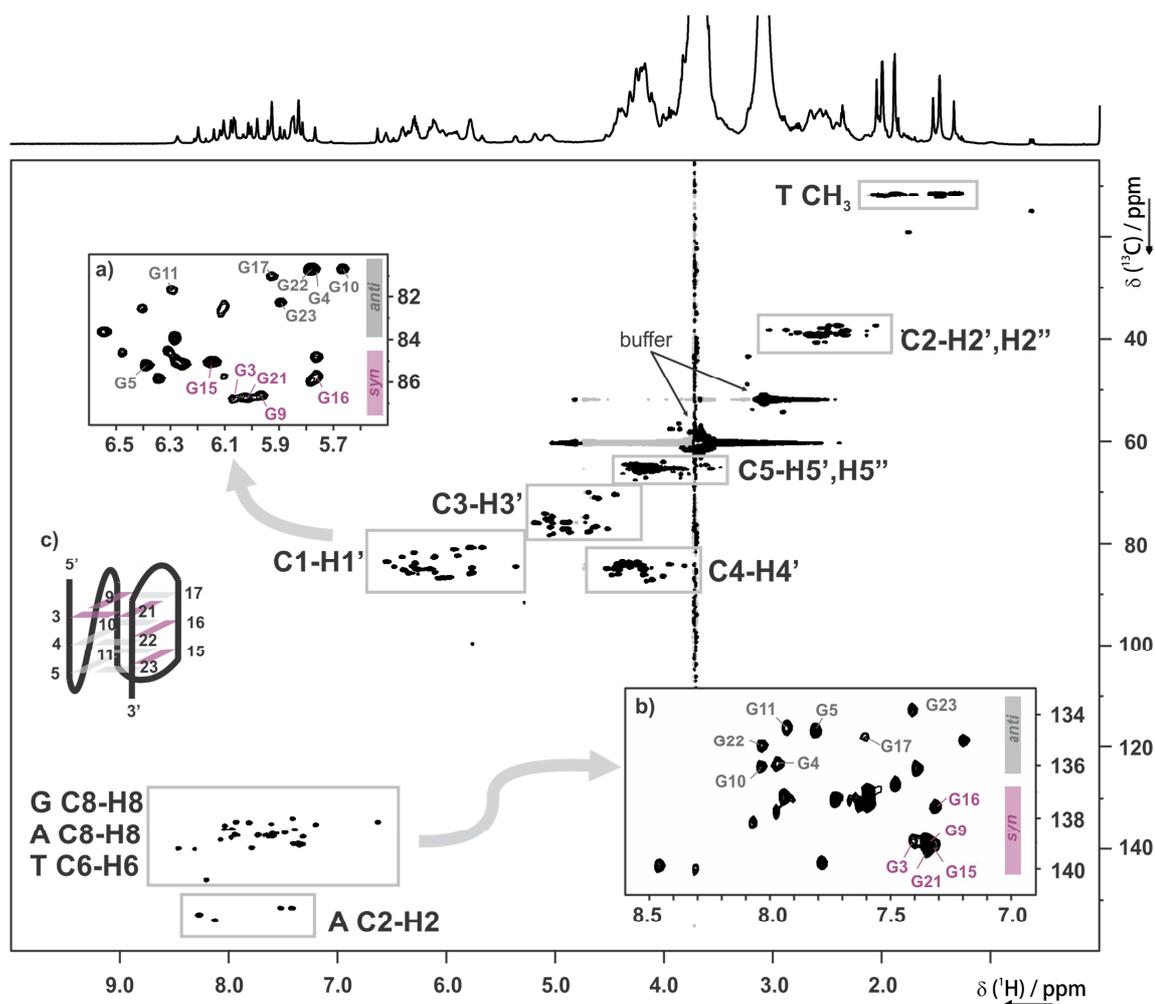


Figure 33 $^1\text{H},^{13}\text{C}$ -HSQC recorded on a natural abundance telomeric DNA (Tel24, d(TTGGGTTAGGGTTAGGGTTAGGA) sample 0.5 mM in 25 mM bis-tris buffer (pH 7.0) containing 70 mM KCl, 100% D₂O, at 600 MHz, 298 K. Zoom on the a) C1'-H1' and b) C8-H8 region with assignment of the guanine residues according to the numbering reported on the structure displayed in c). Anti guanines residues are colored in gray, while syn guanines are colored in violet.

Beyond providing through-bond connectivities necessary for the assignment, the J-couplings provide information on the torsion angles essential for structure determination. An overview of the coupling constants reported for the guanine nucleobase is displayed in **Figure 34a**, while **Figure 34b** summarizes the homo- and heteronuclear ^1J -couplings reported for the deoxyribose sugar and the J_{CP} -couplings between the sugar and the phosphate backbone.

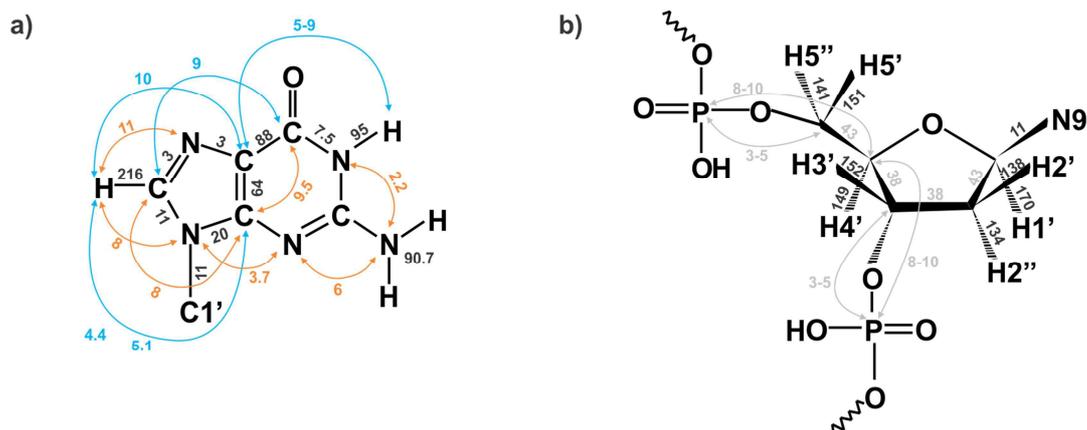


Figure 34 Overview of the coupling constants (Hz) for guanine base and deoxyribose. All the coupling constants in the guanine base (left) were determined on 5'-GMP by Ippel *et al.*^[319] 3J , 2J and 1J are respectively black, orange and cyan. The $^1J_{CC}$ in the sugar moiety (right) were determined for ribose, therefore are indicated in gray. The $^1J_{CH}$ were determined by Ippel *et al.* on the deoxy moiety of the cyclic nucleotide r<pGp(dG)> with 83% N-type (C3'-endo) state populated.^[319] The calculations by Vokáčová *et al.* suggested however a different dependency of $^1J_{C2'H2'}$ and $^1J_{C2''H2''}$ on the sugar pucker ($^1J_{C2'H2'}$ 132.4-137.2 Hz and $^1J_{C2''H2''}$ 123.7-135.2 Hz for C2'-endo, $^1J_{C2'H2'}$ 124.7-128.5 Hz and $^1J_{C2''H2''}$ 133.0-140.7 Hz for C3'-endo).^[320] The $^2J_{CP}$ and $^3J_{CP}$ (gray) have been determined by Schwalbe *et al.* for RNA and GMP, respectively.^[321]

A detailed description of the J-couplings in nucleic acids can be found in the reviews by Ippel *et al.* and Wijmenga *et al.*^[319,322] Selected examples of 3J couplings dependence on torsional angles are shown in **Figure 35**.

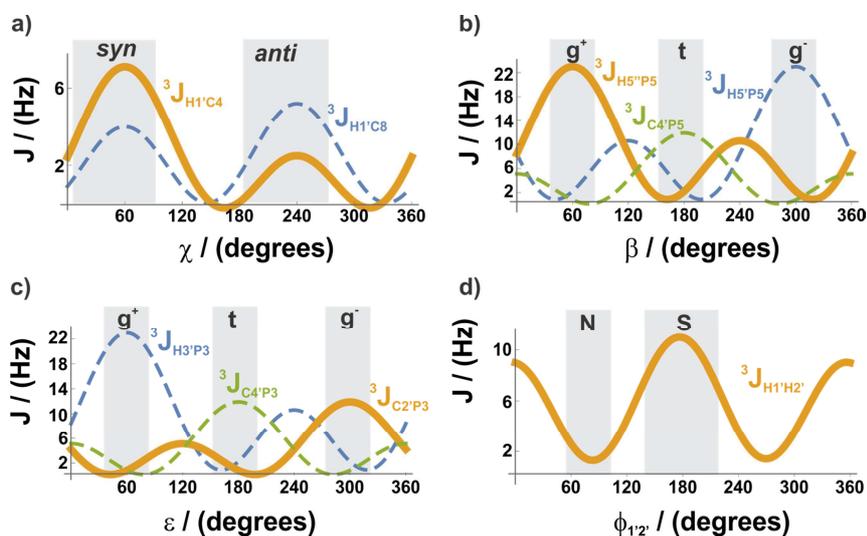


Figure 35 a) Dependence of $^3J_{C8H1'}$ and $^3J_{C4H1'}$ couplings on the glycosidic torsional angle χ according to the Karplus equation parametrized by Ippel *et al.*^[319] b) Dependence of $^3J_{H5'P5}$, $^3J_{H5''P5}$ and $^3J_{C4'P5}$ on the torsional angle β according to the Karplus equation parametrized by Mooren *et al.*^[323] c) Dependence of $^3J_{H3'P3}$, $^3J_{C4'P3}$ and $^3J_{C2'P3}$ on the torsional angle ϵ according to the Karplus equation parametrized by Mooren *et al.*^[323] d) Dependence of $^3J_{H1'H2'}$ on the torsional angle defined by H1'-C1'-C2'-H2' ($\phi_{1'2'}$) according to the generalized Karplus equation proposed by Haasnoot *et al.*^[324] The observed ranges for the torsion angles are highlighted with gray boxes and the corresponding geometry descriptors ($g^+ = +60^\circ$, $g^- = -60^\circ$ and $t = 180^\circ$; *syn*, *anti*, south (S) and north (N) conformations are defined in **Figure 2**) are indicated with black labels.

5.3 Assignment strategies

Assignment of unlabelled DNA oligonucleotides relies mostly on the through-space NOE connectivities. The extreme conformational diversity of the G-quadruplex structures (see Chapter 2.3) determines a broad range of NOE-observable ^1H - ^1H distances (see **Table 6** for comparison with A- and B-DNA duplexes).

The imino region of the NOESY spectrum is shown in **Figure 36a** while panel b displays as example all the imino-imino NOE contacts detected for residue G10: intra-tetrad connectivities (G4 H1-G10 H1, 4.4 Å), sequential inter-tetrad connectivities (G11 H1-G10 H1, 3.8 Å) and non-sequential, inter-tetrad connectivities (G17 H1-G10 H1, 2.9 Å). The imino-aromatic region **Figure 36e** contains valuable information about the folding topology (**Figure 36c**). In fact, the average distance between the aromatic proton H8 and the imino proton H1 belonging to adjacent guanines in the same G-tetrad is 5 Å, therefore H1-H8 interactions (red arrows, **Figure 36d**) are detectable in this region. Assignment of the intra-tetrad connectivities (**Figure 36f**) allows the complete determination of the folding topology. Furthermore, residues belonging to the loops or to the flanking regions that are stacking onto the external quartet may give intense cross-peaks in this region. For example, the stacking of adenine A20 onto the 5'-end tetrad can be inferred from the strong cross-peaks detectable between A20 H2 and all the guanine imino protons (**Figure 36g**). If the G-quadruplex forming oligonucleotide contains thymine residues, further information on the geometry of the capping structure can be obtained from the analysis of the imino-methyl region (**Figure 36h**). **Figure 36i** shows as example the NOE interaction detected between G5 H1 and the methyl group of residue T13.

		Distance (Å)			
		A-DNA [#]	B-DNA [#]	G-quadruplex	
G H1-G H1	Inter-strand	3.8	3.4	3.8	
	Sequential	4.3	3.4	4.4 (intra-tetrad) 2.9 (inter-tetrad)	
G H1'-G H8	Intra-residue	<i>anti</i>	3.8	3.8	3.9
		<i>syn</i>	n.d.	n.d.	2.6
	Sequential	5'- <i>anti-anti</i> -3'	4.6	3.6	3.6
		5'- <i>syn-anti</i> -3'	n.d.	n.d.	4.0
		5'- <i>anti-syn</i> -3'	n.d.	n.d.	> 5.0 *
		5'- <i>syn-syn</i> -3'	n.d.	n.d.	> 5.0
		3'- <i>anti-anti</i> -5'	> 5.0	> 5.0	> 5.0
		3'- <i>anti-syn</i> -5'	n.d.	n.d.	3.4
3'- <i>syn-anti</i> -5'	n.d.	n.d.	> 5.0 *		
3'- <i>syn-syn</i> -5'	n.d.	n.d.	5.0		

Table 6 Imino-imino and anomeric-aromatic ^1H - ^1H distances (Å) measured in regular A-DNA and B-DNA ([#]data from reference [325]. n.d.: not determined) and DNA G-quadruplex (data measured on the PDB structure 2GKU, with the exception* of the data for the 5'-*anti-syn*-3' step, measured on the PDB structure 143D).

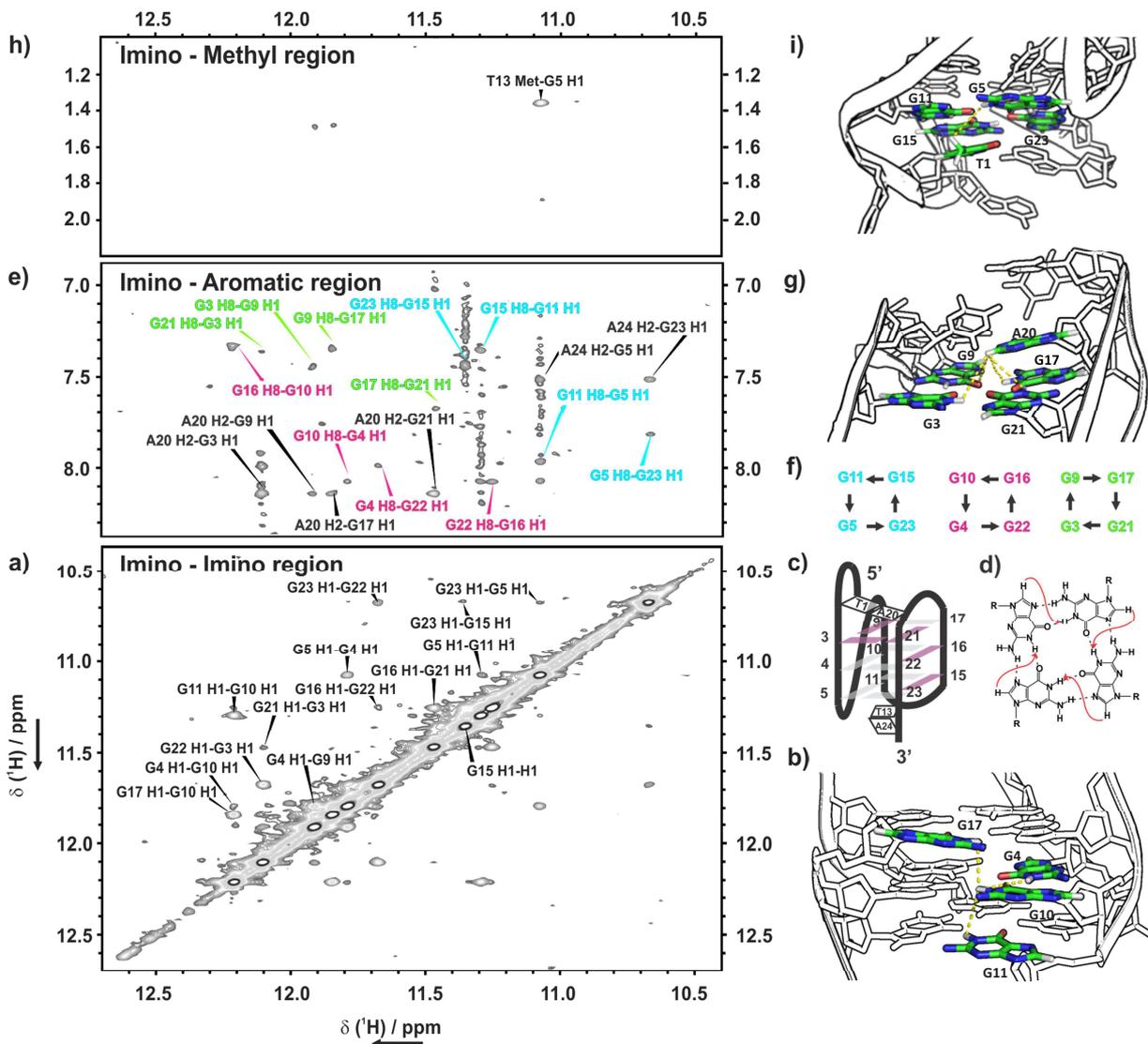


Figure 36 Assignment of the imino region of the NOESY spectrum of 1 mM Tel24 in 25 mM Bis-Tris buffer at pH 7.0 containing 70 mM KCl in 90% H₂O / 10% D₂O, recorded at 600 MHz and 288 K. a) Imino-imino region with b) NOE contacts observed for residue G10 indicated on the NMR structure (PDB: 2KGU) determined by Phan and co-workers.^[304] c) Scheme of the hybrid-1 conformation adopted by Tel24.^[304] d) Scheme of a G-quartet with red arrows indicating the H8-H1 NOE interactions between adjacent guanines expected in the imino-aromatic region shown in panel e). f) Intra-tetrad H8-H1 interactions, with color code according to the labels in panel e). g) NOE contacts between A20 H2 and the 5'-end tetrad observed in the imino-aromatic region (e). h) Imino-methyl region with NOE interaction detected between the methyl group of residue T13 and the imino proton of residue G5 is indicated in panel i).

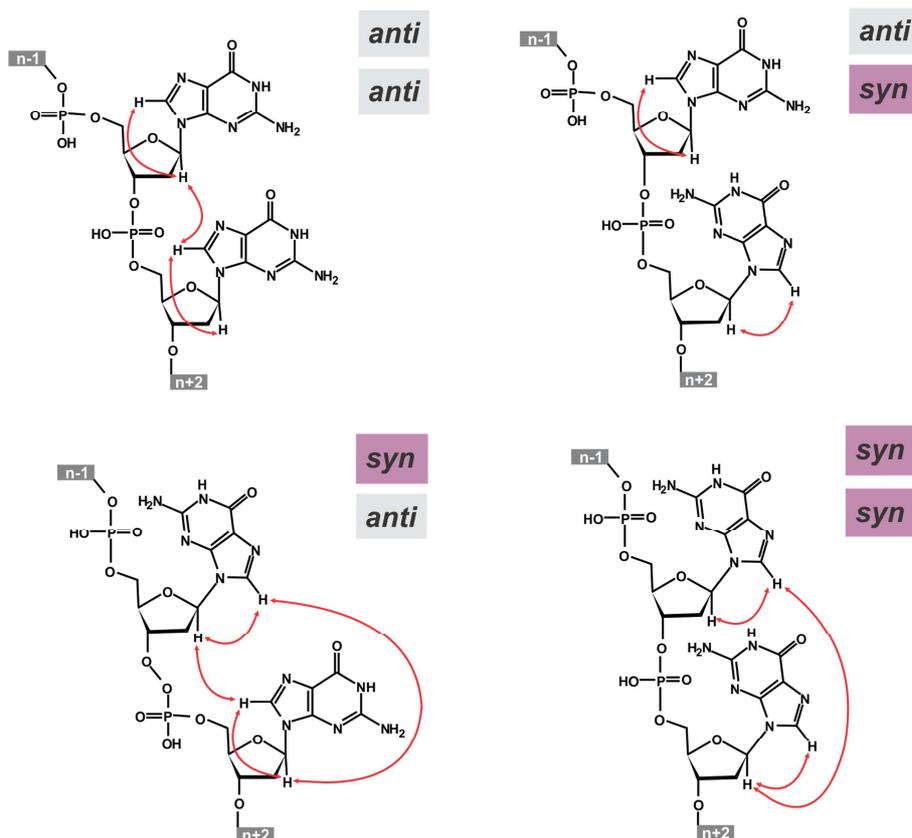


Figure 37 Scheme of the detectable aromatic-anomeric NOE connections (red arrows) according to the relative glycosidic torsion angle conformation of two successive guanine residues. *anti* residues are represented in gray, while *syn* residues in magenta.

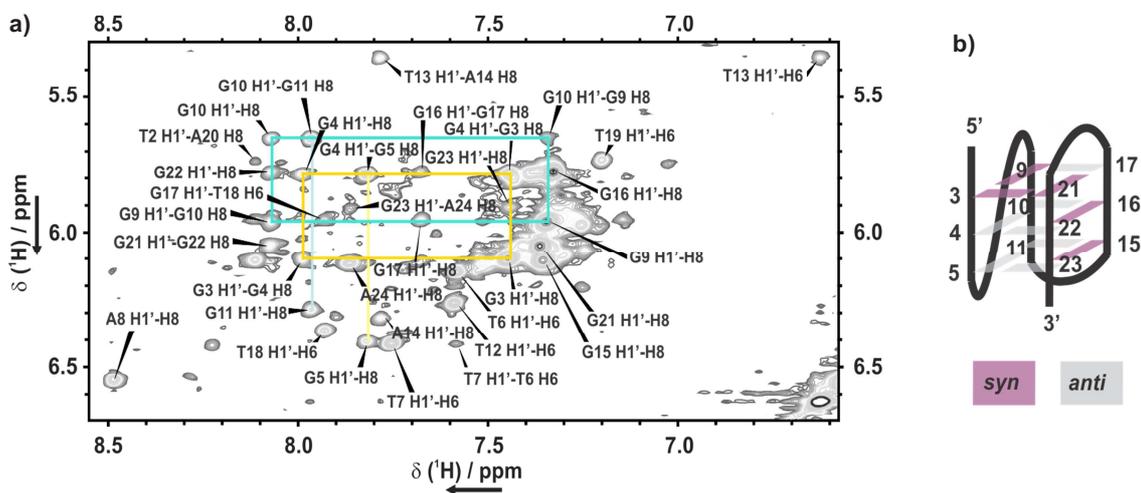


Figure 38 a) Assignment of the aromatic-anomeric region of the NOESY spectrum of 1 mM Tel24 in 25 mM Bis-Tris buffer at pH 7.0 containing 70 mM KCl in 90% H₂O / 10% D₂O, recorded at 600 MHz and 288 K. The H8(n)-H1'(n)-H8(n+1) sequential walk is indicated as example for G3-G4-G5 and G9-G10-G11 in yellow and cyan, respectively. b) Scheme of the hybrid-1 conformation adopted by Tel24^[304] with *anti* and *syn* guanine residues represented in gray and magenta, respectively.

It is possible to trace the H8(n)-H1'(n)-H8(n+1) NOE sequential walk, usually employed for assignment of B-helix DNA and A-helix RNA, also for DNA G-quadruplexes. However, due to geometrical reasons, the sequential pathway is broken at the *anti-syn* step and at the *syn-syn* step

(**Figure 37**). Furthermore, at a *syn-anti* step the cross peak between H8(n) and H1'(n+1) is also detectable, resulting in a squared NOE path. The intra-residue H1'-H8 cross peak for residues in *syn* conformation is usually very strong (H1'-H8' distance = 2.5-2.7 Å). The sequential connectivity in the loop regions may be lost, depending on the loop length and geometry. The assignment of the aromatic-anomeric region of Tel24 is shown in **Figure 38a**. As example, the sequential walk for the G-stretches G3-G4-G5 and G9-G10-G11 is indicated with full lines and the squared NOE connectivity pattern can be detected at the steps G3(*syn*)-G4(*anti*) and G9(*syn*)-G10(*anti*).

Similar NOE sequential walks can be traced in the H8-H2'/H2'' region.

The sugar proton assignment can be confirmed/completed *via* experiments based on through-bond interactions, such as 2D $^1\text{H}, ^1\text{H}$ -TOCSY, $^1\text{H}, ^1\text{H}$ -2D DQF-COSY and 2D $^1\text{H}, ^{31}\text{P}$ -correlation spectroscopy.

The long-range $^3J_{\text{HC}}$ -couplings can be also exploited to correlate *via* the carbon C5 the imino and the aromatic proton belonging to the same guanine residue (see **Figure 39** for scheme of the magnetization transfer). The $^{13}\text{C}, ^1\text{H}$ -HMBC at natural abundance with jump-return water suppression developed by Phan has been widely used to cross-check the assignment of imino and aromatic protons in G-quadruplexes.^[326]

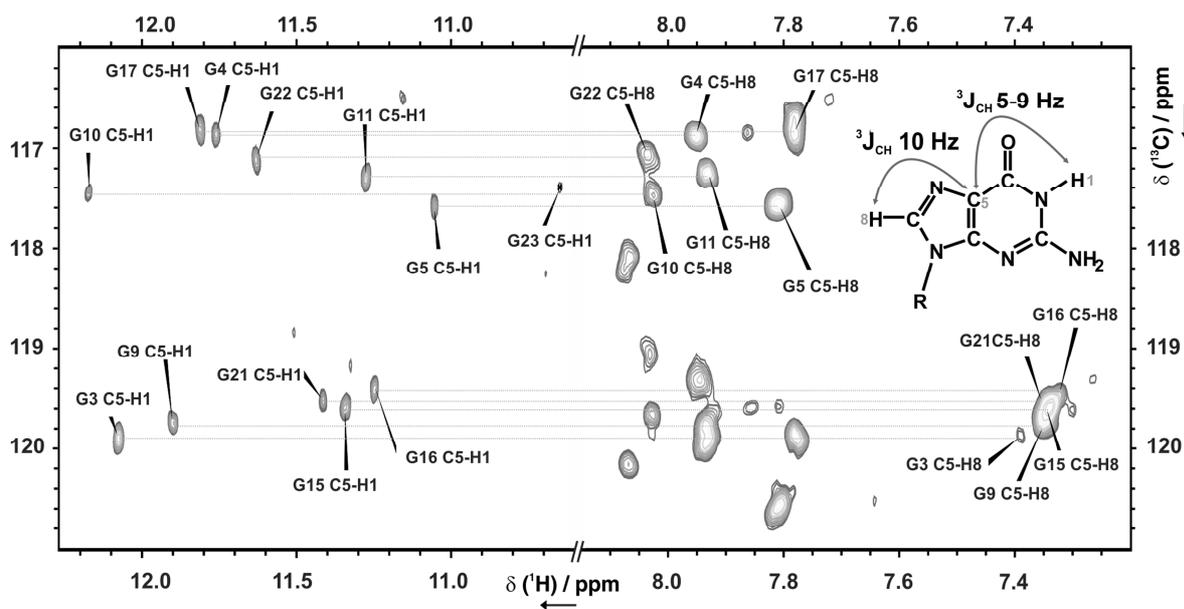


Figure 39 C5 region of the $^{13}\text{C}, ^1\text{H}$ -HMBC recorded on a 2 mM Tel24 sample at natural abundance. J-couplings exploited for the long-range transfer are indicated on the guanine structure. Further experimental conditions: 25 mM Bis-Tris buffer at pH 7.0 containing 70 mM KCl in 90% H_2O / 10% D_2O , recorded at 600 MHz and 298 K. The delay for the long-range magnetization transfer was set to 29 ms.

The assignment of a G-quadruplex can be facilitated if the sample is isotopically enriched. In particular, not all the through-bond correlations can be obtained at natural abundance. Thus, unambiguous assignment *via* site-specific labelling may be necessary in G-rich sequences that are intrinsically characterized by heavy overlap.

The intra-tetrad connectivities, which ultimately provide the folding topology, can be determined on a uniformly ^{15}N -labelled sample using HNN-COSY experiments. The $^2J_{\text{NN}}$ scalar coupling

(6-8 Hz)^[327] can be used to correlate N2 and N7 involved in the G-tetrad Hoogsteen base pairing. The N2-N7 correlation can be established *via* the amino hydrogen H2 or the aromatic proton H8.^[328] This strategy has been applied by Phan *et al.* to determine the structure of the RNA G-quadruplex formed by the 36 nt-long aptamer binding to the fragile X mental retardation protein (FMRP).^[329]

While isotopically enriched RNA is routinely prepared in mg amounts *via in vitro* transcription using T7 polymerase,^[330] comparably cost-effective methods for the synthesis of (large) ¹³C, ¹⁵N labelled DNA are still missing. Enzymatic synthesis of isotopically labelled DNA has been introduced in 1995 by Zimmer and Crothers.^[331,332] Later, an *in vivo* method, based on the amplification of plasmids in *E. coli* cells grown on labelled minimal medium, and an *in vitro* PCR-based method were proposed by Louis *et al.* for preparation of isotopically labelled dsDNA.^[333] A new method employing self-primed PCR with asymmetric endonuclease digestion has been recently developed by Nelissen *et al.* for preparation of isotopically labelled ssDNA.^[334] However, chemical synthesis remains the most versatile method to produce labelled DNA and it allows to obtain uniformly or site-specifically labelled material with a variety of labelling schemes.^[335] The main drawback of this strategy is the high cost of the isotopically labelled phosphoramidites. NOE-based assignment of G-quadruplexes can be ambiguous due to the overlap of the signals and to the fact that imino-imino inter-tetrad cross-peaks may arise from intra-strand sequential interactions or inter-strand interactions (**Figure 36**). ¹³C/¹⁵N site-specific labelling may be necessary for unambiguous assignment. A relatively affordable approach was introduced in 2002 by Phan and Patel, showing that diluting the labelled material (0.5 - 4.0% enrichment level in 2-4 mM DNA samples) allows unambiguous assignment.^[336] For example, a 1% ¹³C, ¹⁵N enriched sample exhibits a 4-fold increase of ¹⁵N (abundance 1.37%) and a 2-fold increase of ¹³C (abundance 2.1%) compared to the natural abundance (0.37% ¹⁵N and 1.1% ¹³C). Since then, the use of site-specifically labelled samples at low enrichment levels has become very popular and it is routinely used to unambiguously assign the guanine imino protons, the aromatic protons and the sugar protons.^[120,217,329,337]

Isotope-filtered experiments allow to discriminate between protons bound to ¹²C/¹⁴N and protons bound to ¹³C/¹⁵N.^[338] **Figure 40** summarizes the assignment strategy for residue G17 in Tel24 sequence, which has been used in the folding kinetic experiments presented in Chapter 7. ¹⁵N-filtered 1D ¹H and ¹H, ¹³C-HSQC spectra recorded on a 150 μ M DNA sample ¹³C, ¹⁵N site-specifically labelled at position G17 (98% enrichment level) allowed to unambiguously assign H1 and H8 protons belonging to G17 in each conformation populated during the folding process.

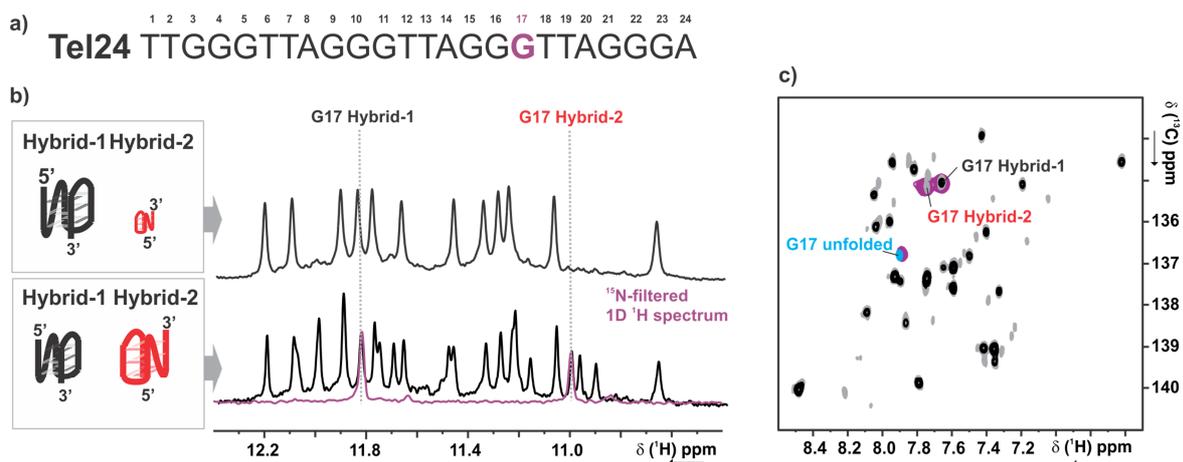


Figure 40 Assignment of imino and aromatic proton of residue G17 (violet) of telomeric sequence Tel24 (a) in hybrid-1 (black schematic) and hybrid-2 (red schematic) conformation using DNA ^{13}C , ^{15}N site-specifically labelled at position 17 (violet; 98% isotopically enriched, 150 μM DNA). b) Bottom: overlay of the imino region of the 1D ^1H spectrum (black) at the beginning of the folding kinetic, when the two conformations are equally populated, with the corresponding ^{15}N -filtered spectrum (violet). Top: imino region of the 1D ^1H spectrum at the end of the folding kinetic, when the hybrid-1 conformation is predominantly populated. c) Overlay of the aromatic region of ^1H , ^{13}C -HSQC at different stages of the folding kinetics. The cyan spectrum was recorded on the ^{13}C , ^{15}N sample labelled at position 17 before inducing the folding (only the unfolded state is populated) and the violet spectrum after inducing the folding (hybrid-1, hybrid-2 and unfolded states are populated). The gray spectrum was recorded on a sample at natural abundance after inducing the folding, while the black spectrum at the end of the kinetic. See Chapter 7.1 for further details.

5.4 Binding and dynamics of cations studied by NMR

The thermodynamics of cation binding as well as the dynamics of the cations movement within the channel can be investigated by NMR.

NMR was used to explain the selectivity of G-quadruplexes for K^+ versus Na^+ . Analysis of the ^1H chemical shift resulting from replacement of Na^+ with K^+ in the bimolecular quadruplex formed by $[\text{d}(\text{G}_3\text{T}_4\text{G}_3)]_2$ allowed to conclude that the preference for Na^+ over K^+ is driven by the difference in cation dehydration energy and not by the difference in coordination energy.^[65]

DFT calculations showed that the $^2J_{\text{NN}}$ couplings correlating N2 and N7 are connected to the hydrogen bond geometry in the G-tetrad, *i.e.*, to the hydrogen bond distances.^[339] The dependence of $^2J_{\text{NN}}$ couplings on the temperature and nature of the stabilizing cation has been studied by Dingley *et al.* on the sequence $[\text{d}(\text{G}_4\text{T}_4\text{G}_4)]_2$, derived from *Oxytricha nova* telomeric DNA.^[66] The J-coupling analysis highlighted details that were not picked up by X-ray crystallography or NMR structures based on NOE restraints: for example, that the $\text{N2}\cdots\text{H2}\cdots\text{N7}$ at the 5'-end are the most thermolabile hydrogen bonds.

Heteronuclear liquid state NMR has also been employed to directly detect the metal ion binding. Hud *et al.* showed already in the late Nineties that $^{15}\text{NH}_4^+$ is a valuable spin- $\frac{1}{2}$ probe for cation binding in solution state NMR.^[340] In presence of $^{15}\text{NH}_4^+$, the DNA sequence $[\text{d}(\text{G}_4\text{T}_4\text{G}_4)]_2$ forms a symmetric, diagonally looped, antiparallel dimeric quadruplex with four G-tetrads, *i.e.*, the same folding topology as in a Na^+ -containing solution.^[341] Three $^{15}\text{NH}_4^+$ cations lodged between the G-tetrads were directly detected via ^{15}N -filtered experiments and, in contrast with the X-ray

structure reported in presence of K^+ ,^[342] no $^{15}NH_4^+$ were observed in the poly-thymine loops. N_z chemical exchange heteronuclear single-quantum coherence (N_z Ex-HSQC) experiments at different mixing times allowed to determine for the $^{15}NH_4^+$ in the inner site a rate of exchange with the $^{15}NH_4^+$ in the outer site of 4 s^{-1} at 283 K. More recently, Plavec and co-workers have used this approach to elucidate the mechanism of movement of cations in the quadruplex channel (see Chapter 2.1 for more details).^[67–73] Although $^{15}NH_4^+$ has been shown to be a valuable replacement of alkali ions in liquid NMR studies, it is important to keep in mind that ammonium ions are potential hydrogen donors and have specific pH requirements.

$^{205}Tl^+$ has also been used in solution state NMR as spin- $\frac{1}{2}$ probe for investigation of cation binding to G-quadruplexes.^[51,343] $^{205}Tl^+$ has a high natural abundance (70.5%), a relative sensitivity to proton of 0.13 and chemical properties similar to K^+ (e.g., the ionic radius, see **Table 1**). Gill *et al.* have used 1H - ^{205}Tl scalar couplings and direct detection of ^{205}Tl to localize Tl^+ within the coordination sites and characterized Tl^+ -bound form of $[d(G_4T_4G_4)]_2$ (PDB code: 1K4X).^[51]

While solution state NMR was well established for spin- $\frac{1}{2}$ probes, the investigation of quadrupolar nuclei such ^{23}Na and ^{39}K or ^{87}Rb has been more challenging.

Early studies from Laszlo and co-workers reported the direct detection of dehydrated $^{23}Na^+$ bound to guanosine self-assemblies.^[44,45] Braunlin and co-workers proposed twenty years ago that the alkali metal cations ^{23}Na and ^{39}K bound to the G-quadruplex core result in an extreme line broadening that makes them “invisible” to solution-state NMR because of very rapid quadrupolar relaxation.^[344] More recently, Wu and co-workers developed a new approach which allows obtaining high-resolution spectra of quadrupolar nuclei.^[345,346] In particular, they can directly detect three classes of cations: unspecifically bound to the phosphate backbone, specifically bound to the G-quadruplex channel and bound to the loop regions.

It has to be mentioned that solid state NMR has been also employed to directly detect the cations bound to G-quadruplexes. Solid state ^{23}Na NMR studies on G-quadruplex DNA were reported first by Griffin and co-workers and then by Wu and co-workers.^[347,348] The binding of potassium cations to G-quadruplex structures was also directly detected by solid state ^{39}K NMR, using a magnetic field of 19.5 T to overcome difficulties due to the low γ of ^{39}K .^[349] Furthermore, ^{17}O NMR spectroscopy was used to probe the interaction between cations and carbonyl groups involved in G-tetrad formation.^[350]

5.5 Real-time NMR to monitor G-quadruplex folding

NMR is a powerful technique to investigate the dynamics of biomacromolecules at atomic resolution. Several experimental techniques covering different time scales are available nowadays to investigate the various molecular motions and the dynamic processes (see **Figure 41a**). More about NMR solution-state methods to characterize the dynamics of nucleic acids can be found in the review [351] by Al-Hashimi.

Real-time NMR is optimally suited to investigate processes that are occurring on a time scale longer than 1 s, such as folding and conformational transitions (**Figure 41a**, black).

The folding can be triggered by disturbing the chemical equilibrium directly in the spectrometer in several ways, *e.g.*, illumination of photo-caged or photo-sensitive molecules,^[352] rapid change of the temperature^[353] or the sample composition (pH, concentration of cations or denaturing agents).^[354]

We have employed the latter approach to study the folding kinetics of telomeric G-quadruplex (Chapter 7.1), using the rapid mixing device shown in **Figure 41b**^[355] which allows to trigger the folding by changing abruptly the K^+ concentration in solution. An insert, connected to a pneumatic injection syringe situated outside the magnet, is placed inside a Shigemi NMR tube. With this set-up, 300 μ L of unfolded DNA in a K^+ -free buffer can be mixed with 40 μ L of KCl solution trapped between two air bubbles in the insert. The NMR tube serves as mixing chamber. The set-up is optimized to guarantee a homogeneous mixing of the solutions inside the tube. The dead time (1-2 s) is basically depending on the repetition delay and DNA concentration. After a single injection event a series of consecutively acquired 1D 1H NMR spectra is recorded as pseudo 2D experiment (**Figure 41c**). The kinetic information can be extracted by plotting the intensity of selected peaks as a function of time.

In case of high degree of overlap in the 1D 1H NMR spectrum, provided that the sample is labeled, a variety of fast NMR techniques can be used to monitor the kinetics of folding. However, the examples of fast real-time multidimensional NMR applied to nucleic acids are rare in the literature.^[356] For an overview of the recent progress in this field, see the review [357] and references therein.

NMR of G-quadruplexes

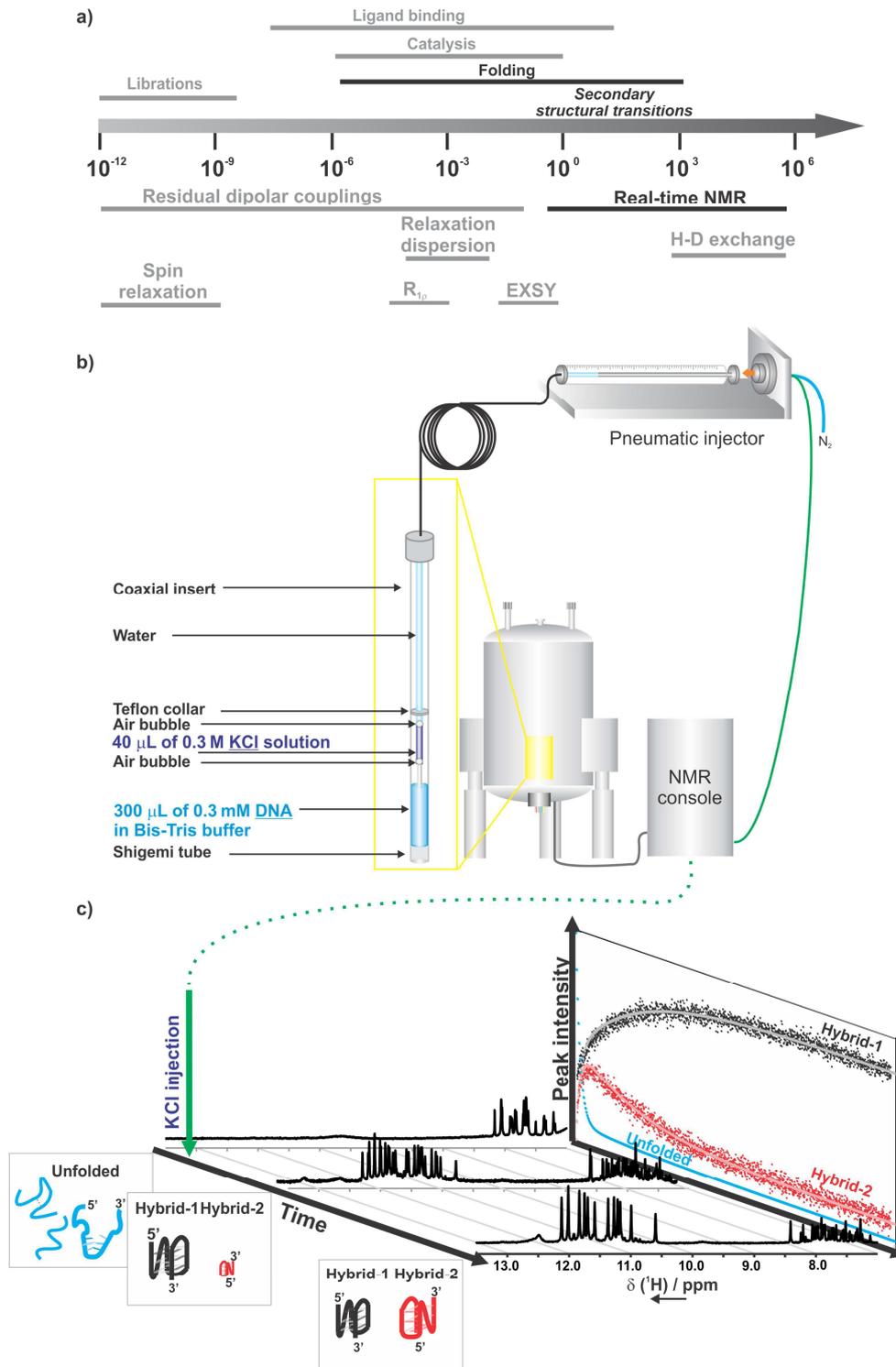


Figure 41 a) Time scales of the different molecular motions/dynamic processes and NMR methods available to investigate the dynamics on the different time scales. The processes that can be investigated with real-time NMR are colored in black. b) Scheme of the experimental setup for real-time NMR to monitor the folding of G-quadruplex with zoom on the rapid mixing device developed by Mok *et al.*^[355] c) The strategy used to monitor the folding kinetic of the telomeric sequence Tel24 (see Chapter 7.1 for the discussion of the details of the folding mechanism). A series of 1D ^1H NMR spectra is recorded as pseudo-2D experiment after inducing the folding with the injection of KCl triggered by the pulse program.

6 G-quadruplex ligands

This chapter gives a short overview of the different classes of G-quadruplex ligands and their possible binding modes and discusses the effective biological relevance and drug-likeness of the G-quadruplex ligands developed until now. Small molecules targeting G-quadruplex have been surveyed in many reviews with different perspectives by several groups.^[264,358–361] An online database for G-quadruplex ligands (www.g4ldb.org) is also available and contains up to date over 900 hits.^[362]

Structural features of G-quadruplex ligands

The classical features of a G-quadruplex ligand are an extended flat aromatic moiety able to form π - π interactions with a G-tetrad and the presence of positively charged groups able to bind the negatively charged backbone *via* electrostatic interactions.

The aromatic moiety should in theory be much larger than the one of a classical duplex intercalator (*e.g.*, ethidium bromide, proflavine...) in order to increase the duplex/quadruplex selectivity and maximize the stacking interaction with the quartet surface. The positive charges may arise from *in situ* protonation of an amine group, *via* N-methylation of aza-aromatic groups or *via* metal complexation. Besides using the cationic nature criterion, G-quadruplex ligands can also be classified from a structural point of view in two groups: those featuring a polycyclic aromatic core and those containing not-fused hetero-aromatic compounds, which are frequently, but not always, embedded in polycyclic structures. A very popular approach adopted in the G-quadruplex ligands design is based on functionalizing the aromatic moiety with positively charged side chain(s), with the aim to “clamp” the quadruplex core. In this way the hydrophobic aromatic core is retained, and at the same time the protonated side arms enhance the water solubility. Although the principles guiding the design of most of the ligands reported until now assume a stacking binding mode to the G-quartet, a novel binding mode *via* groove interaction has been reported for distamycin.^[363] The same binding mode to G-quadruplexes was observed for non-planar compounds without aromatic rings and for a neutral derivative of distamycin.^[364,365]

Figure 42 shows a selection of representative G-quadruplex binding ligands that have been reported to have a G-quadruplex-related biological activity.

The very early studies from the labs of Neidle and Hurley were focussed on amido-anthraquinone derivatives.^[267] However, due to the low water solubility and scarce duplex vs quadruplex selectivity, the amido-anthraquinone core was gradually replaced with an acridine core. In particular the quadruplex/duplex affinity and the potency of telomerase inhibition of 3,6,9-trisubstituted acridines (BRACO-19, **Figure 42a**) revealed to be much higher than for the corresponding 3,6-disubstituted acridines.^[268,366] The crystal structure of BRACO-19 with the bimolecular telomeric quadruplex [d(TAGGGTTAGGGT)]₂ (PDB: 3CE5)^[367] indeed confirmed the rationale behind the ligand design. With the aim to increase the degree of π - π stacking with an external quartet, the acridine core has also been extended to generate RHPS4 (**Figure 42b**),^[368]

containing a pentacyclic moiety with a positive charge but without side-chains. Some details into the binding mode of RHPS4 to the tetrameric quadruplex [d(TTAGGGT)]₄ have been elucidated by NMR (PDB: 1NZM).^[369] **Figure 42c** shows the molecular structure of Quarfloxin (CX-3542), the only G-quadruplex targeting compound that up to date has progressed to Phase II clinical trials (ClinicalTrials.gov Identifier: NCT00780663, see also Chapter 4.1 for possible mechanism of action). The fluoroquinolone scaffold was known to have poisoning activity against topoisomerase II and bacterial gyrase. Hurley and co-workers developed a fluoroquinolone-based ligand with topoisomerase II activity and with G-quadruplex interaction capability.^[370] The lead optimization process was then completed by Cylene Pharmaceuticals and produced the compound quarfloxin, without topoisomerase II activity and with higher quadruplex/duplex selectivity.

Natural compounds have also been reported to bind to the G-quadruplex with different efficiency. In particular, the isoquinoline alkaloid berberine (**Figure 42d**) has frequently been used as lead compound. The interaction studies of berberine and the structurally related alkaloid sanguinarine with G-quadruplex structures derived from human telomeres will be presented in Chapter 6.1. After the publication of this article, the crystal structure of berberine with a telomeric DNA G-quadruplex was reported (PDB: 3RSR, **Table 7** and **Figure 42**).^[371] The natural alkaloid cryptolepine served as scaffold to develop the quindoline compound displayed in **Figure 42e**, which showed the ability to stabilize the *c-MYC* G-quadruplex and inhibit the *c-MYC* expression in cancer cells. The NMR structure of the *c-MYC* G-quadruplex with ligand e) has been solved by NMR by Yang and co-workers (PDB: 2L7V, **Table 7** and **Figure 43**).

The naphthalene diimide moiety, known to bind to duplex DNA, was modified in Neidle's group to obtain tri- and tetra-substituted analogues with improved selectivity for quadruplex over duplex.^[372] On the basis of biophysical and biological data, the tetra-substituted derivatives were further developed and various derivatives differing in side chain length and nature of the functional groups were synthesized (**Figure 42f-g**) following a structure-based approach (**Table 7** and **Figure 43**).^[373-375] Interestingly, the large planar surface featuring perylene diimide derivatives (see as example the compound PIPER, **Figure 42h**)^[376] does not guarantee better G-quadruplex over duplex selectivity but impairs the hydrosolubility of the compound. On the other side, the quadruplex selectivity of perylene diimide derivatives seems to increase under experimental conditions promoting ligand self-aggregation.^[377]

The groups of Mergny and Fichou reported on the use of bisquinolinium derivatives as scaffold for G-quadruplex ligands. In particular, the phenantholine analogue Phen-DC3 (**Figure 42i**) showed exceptional quadruplex/duplex selectivity and high capability of enhancing quadruplex thermal stability.^[378] The aromatic surface of Phen-DC3 was designed to perfectly match the aromatic area of a G-quartet. In fact, the NMR structure of this ligand in complex with a unimolecular quadruplex (PDB: 2MGN, **Table 7** and **Figure 43**) reveals an optimal overlap between the tetrad and the phenanthroline and quinolinium groups.^[216]

Carbazole derivatives have also been shown to bind to G-quadruplexes. The fluorescent carbazole moiety was employed for the first time by Chang and coworkers more than one decade ago^[379] and has been engineered to drive conformational transitions of G-quadruplexes^[130] and to

visualize G-quadruplex structures formed at mitochondrial DNA in living cells.^[155] Dash and co-workers have recently optimized the modular synthesis of various bis-triazolyl carbazole derivatives. The compound shown in **Figure 42j** revealed to be a fluorescent “light-up” probe for selective detection of *c-MYC* quadruplexes over duplex DNA in living cells, with potential ability to down-regulate the *c-MYC* expression in cancer cells. The details of the screening, the biological essays and the NMR characterization of the interaction with the *c-MYC* G-quadruplex will be presented in more detail in the research article in Chapter 6.2.

Despite their low quadruplex/duplex selectivity, the cationic meso-methylpyridinium-substituted porphyrin TMPyP4 and its metal complexes (**Figure 42k-l**) are amongst the most popular ligands used as a reference in the screening of G-quadruplex interacting compounds. Actually, the binding mode of TMPyP4 to G-quadruplexes is controversial. While the NMR structure of the cationic porphyrin TMPyP4 (**Figure 42k**) with the *c-MYC* quadruplex shows external stacking on G-tetrads (PDB: 2A5R, **Table 7** and **Figure 43**), spectroscopic data suggest possible intercalation between the tetrads for the complex between TMPyP4 and G-wires.^[380] Although the N-methyl mesoporphyrine IX (**Figure 42m**) binds to G-quadruplex with a weaker affinity compared to TMPyP4, it displays a much higher quadruplex/duplex selectivity and has been reported to recognize selectively parallel folding topologies.^[381] A crystal structure of the ligand N-methyl mesoporphyrine IX (**Figure 42m**) with telomeric DNA has been recently reported (PDB: **Table 7** and **Figure 43**).^[382]

Several metal-containing ligands have been shown to interact effectively with G-quadruplex DNA, but up to date, structural data are available only for the salphen-derivative (**Figure 42n**) in complex with Cu(II) (PDB: 3QSC) and Ni(II) (PDB: 3QSF).

The compound Pyridostatin (**Figure 42o**) was synthesized in the Balasubramanian lab and is one of the most effective ligand reported until now for stabilizing the G-quadruplex melting temperature with a remarkable selectivity for quadruplex over duplex.^[383] It has also been shown to induce cell apoptosis *via* uncapping of POT1 and to induce DNA damage. The ligand pyridostatin has been employed in landmark studies to identify in the genome G-quadruplex structures.^[152,283,384]

The natural compound telomestatin (**Figure 42p**) was isolated in 2001 from the bacterium *Streptomyces anulatus* and has a polycyclic, neutral scaffold which geometrically matches the G-tetrad surface.^[385] Due to its neutral and cyclic nature, it is not able to interact with the DNA duplex. Furthermore, it is a very potent telomerase inhibitor, able to induce apoptosis of cancer cells *in vitro* and *in vivo*.^[386] Despite its very promising structural and biological properties, telomestatin is barely soluble in water and difficult to prepare in large-scale. Many telomestatin analogues have been synthesized in order to circumvent at least the solubility problem. In particular, the macrocycle L2H2-6OTD, containing two lysine residues (**Figure 42q**), has been structurally characterized by NMR in complex with human telomeric quadruplex (PDB: 2MB3, **Table 7** and **Figure 43**).

Based on the Template-Assembled Synthetic G-Quartet (TASQ) developed in Sherman's group, several “nature-inspired” G-quadruplex ligands bearing guanine/guanosine units have been synthesized.^[48,387,388]

The dansyl-diguanosine derivative DDG (**Figure 42r**) was designed in the Dash lab with the aim to conjugate the biomimetic properties of guanosine units with the biocompatible fluorescent dansyl probe.^[389]

DDG can bind to the *c-MYC* G-quadruplex with high affinity and selectivity and shows an enhancement in fluorescence emission upon G-quadruplex binding, which makes it a potential biomarker for G-quadruplexes in living cells. Its binding mode to the *c-MYC* G-quadruplex has been investigated by NMR and the detailed characterization of DDG will be presented in Chapter 6.3. The ligand DDG can also self-assemble into a G-quartet-based ionic channel that can be incorporated into phospholipid bilayers.^[390]

Interesting examples of multitasking ligands have recently been reported by Monchaud and co-workers. The ligand NaphthoTASQ (**Figure 42s**) features a naphthalene core linked to four guanine units. Upon interaction with a G-quadruplex, the naphthalene templates the self-assembly of the guanine units into a G-quartet, which is able to stack onto the G-quadruplex. The naphthalene fluorescence is quenched in the open form and restored in the G-quadruplex-bound, closed form. NaphthoTASQ is therefore a synthetic G-quartet-based “light-up” probe.^[156]

G-quadruplex ligands

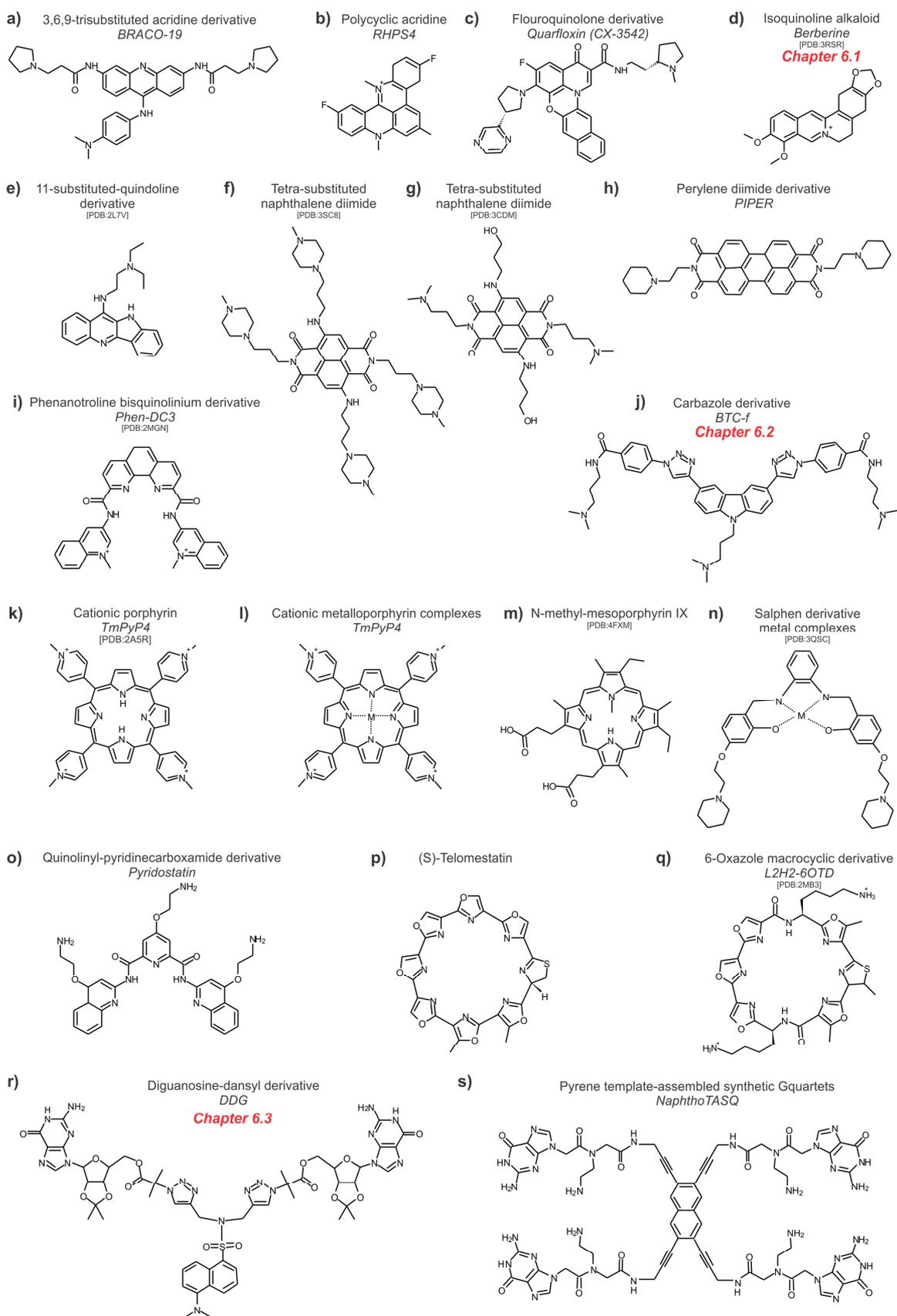


Figure 42 Selected representative G-quadruplex ligands. If available, the common name of the compound is indicated in *italic*, and the PDB code of the structure with G-quadruplex is indicated in brackets.

Methods to investigate the interaction of G-quadruplex with ligands

The interaction of ligands with G-quadruplexes has been investigated using a variety of experimental techniques that have recently been reviewed by Murat *et al.*^[361]

Essentially, the results are strongly dependent on the experimental technique and on the experimental conditions: different laboratories using the same experimental conditions may produce different results. The absolute values of the ligand-induced stabilization of the G-quadruplex melting temperature reported in literature are extremely variable, as well as the binding affinities. Furthermore, the telomerase inhibition potency has frequently been measured using the telomerase repeat amplification protocol (TRAP), composed of a first step where telomerase elongates a substrate and a second PCR step to amplify the product of telomerase elongation. It is now accepted that the ^{tel}IC₅₀ values as determined by the standard TRAP assay are often overestimated due to the ligand-induced inhibition of the PCR step.^[391] Therefore, although general trends can be identified, direct quantitative comparison of literature results is not always meaningful.

Furthermore, it may not be strictly correct to compare the FRET-derived data (melting temperatures and ligands effects on end-to-end distances) directly with the data derived from unlabelled DNA sequences by CD, NMR or UV studies, since the fluorescent labels may interact extensively with the G-quadruplex and perturb the G-quadruplex-ligand interaction.^[392]

The amount of structural data, either from X-ray or NMR, on ligand-G-quadruplex complexes is still rather limited and therefore there are very few examples of structure-based rational design of G-quadruplex ligands (such as the series of naphthalene diimide derivatives developed in Neidle's lab^[372-374,393]).

The ligand-G-quadruplex complexes deposited in the PDB should be analyzed with caution, since there are many tetra- or bi-molecular quadruplexes present that do not possess any loop or somewhat "artificial" loops without physiological relevance. Therefore, only unimolecular quadruplex structures will be considered in the following discussion (**Table 7** and **Figure 43**). Moreover, as discussed in Chapter 2.3, the polymorphism which features the G-quadruplex structures in solution is drastically reduced in crystals. All the crystal structures of telomeric DNA reported up to date are characterized by an all-parallel topology with propeller loops, while in solution this topology has only been observed in presence of polyethylene glycol (see Chapter 2.4). The structural data reported for G-quadruplexes in gene promoter regions are mainly NMR-based and only one crystal structure for the *c-KIT* promoter gene has been reported until now. Interestingly, it is in agreement with the NMR structure (see Chapter 3.2), therefore the discrepancies between X-ray data and NMR data typically observed for telomeric quadruplexes might not be a feature of quadruplexes derived from other genomic locations. This complicates the structure-based design process and suggests that the structural data reported in literature have to be evaluated critically.

On the basis of experimental X-ray, NMR and biophysical data, the most plausible binding mode is stacking on the external 3'- or 5'-end quartet (**Figure 43**), while intercalation between the quartets has never been observed. A survey of the crystal structures of human telomeric

G-quadruplex ligands

quadruplex in complex with small ligands has recently been published by Neidle and co-workers.^[394] All the analyzed structures have the same folding topology (all parallel) but the -TTA- propeller loops reveal a high degree of freedom. The loops can be classified in 12 distinct conformational groups and the most commonly observed loops featured a so-called “TAT intercalation” motif, with the adenine base stacked on top of the 5'-thymine and the second thymine nearly stacked on the external face of the adenine. The Authors conclude that the structures of the loops may be conserved in quadruplexes complexed with structurally similar ligands (e.g., naphthalene diimide derivatives). If possible, *in silico* docking methods should use as target a quadruplex structure from a complex with a structurally similar ligand.

G4 type	G4 topology	Ligand	[DNA]:[Lig.]	Method	PDB code	Ref.
telomeric	All-parallel, propeller	Berberine d)	1:4	X-ray	3R6R	[371]
<i>c-MYC</i>	All-parallel, propeller	Quindoline der. e)	1:2	NMR	2L7V	[217]
telomeric	All-parallel, propeller	Naphth. diimide der. f)	1:1	X-ray	3SC8	[374]
telomeric	All-parallel, propeller	Naphth. diimide der.	1:1	X-ray	3T5E	[374]
telomeric	All-parallel, propeller	Naphth. diimide der. g)	1:4	X-ray	3CDM	[373]
telomeric	All-parallel, propeller	Naphth. diimide der.	1:1	X-ray	3UYH	[375]
telomeric	All-parallel, propeller	Naphth. diimide der.	1:1	X-ray	4DA3	[375]
telomeric	All-parallel, propeller	Naphth. diimide der.	1:1	X-ray	4DAQ	[375]
<i>c-MYC</i>	All-parallel, snapback	Phen-DC3 i)	1:1	NMR	2MGN	[216]
<i>c-MYC</i>	All-parallel, snapback	TMPyP4 k)	1:1	NMR	2A5R	[197]
telomeric	All-parallel, propeller	Mesoporph. der. m)	1:1	X-ray	4FXM	[395]
telomeric	Hybrid-1	Telomestatin der. q)	1:1	NMR	2MB3	[332]

Table 7 NMR and X-ray structures of intramolecular G-quadruplex (G4) structures in complex with ligands deposited in the Protein Data Bank. The structures of the ligands are indicated in **Figure 42** according to the letter code.

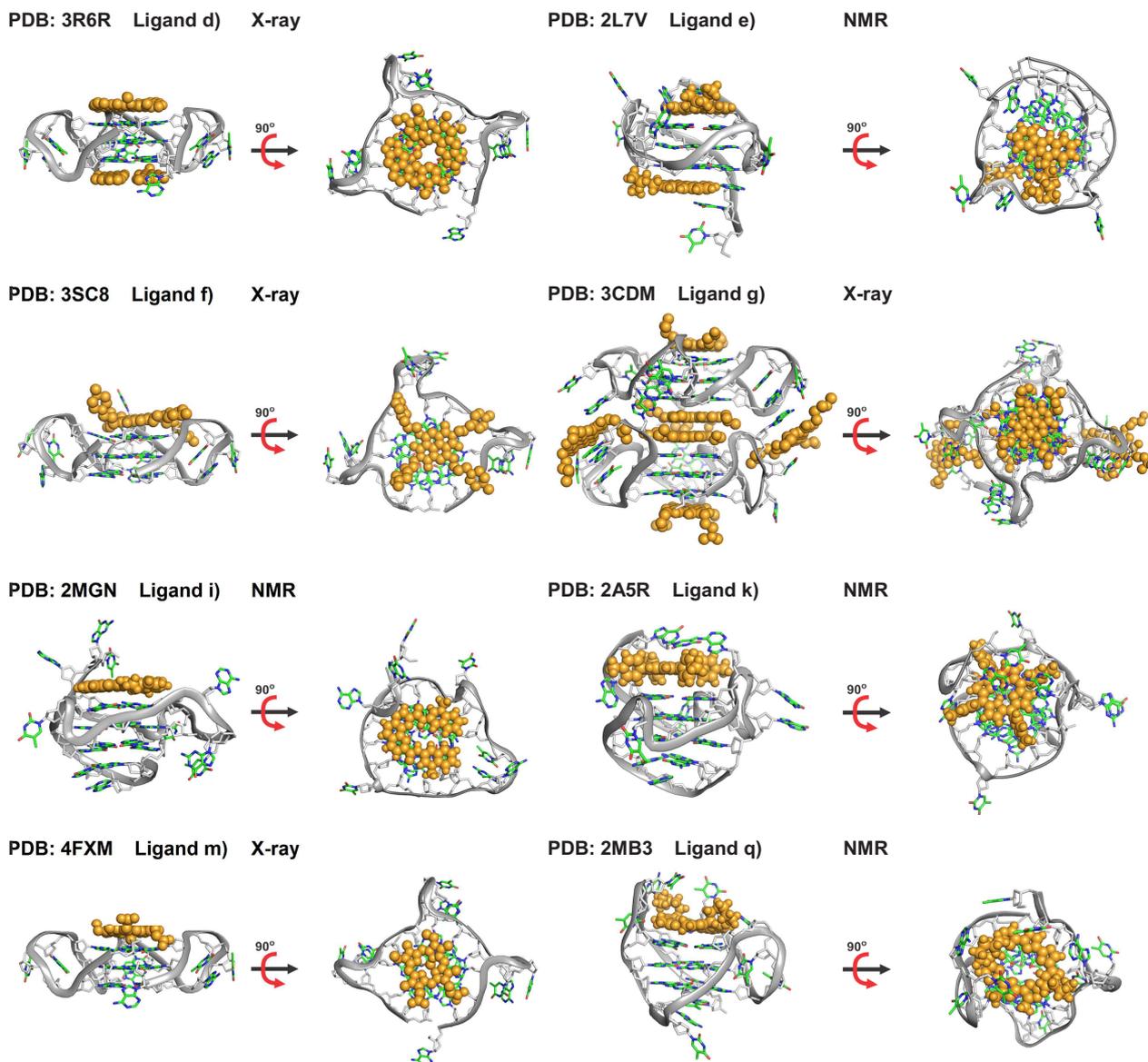


Figure 43 Experimentally determined interaction modes of different G-quadruplex ligands as reported until now by NMR or X-ray. A selection of structures from **Table 7** is shown according to the following criteria: 1) in case of an asymmetric crystal unit with more quadruplexes, they are all shown only if there is a ligand stacking at the interface (3CDM), otherwise half unit is omitted (3SC8); 2) in case of structurally similar ligands (3UYH, 3SC8 and 3T5E), only one representative is shown. The ligands are shown in spheres representation (orange). The structures of the ligands are indicated in **Figure 42** according to the letter code.

All the X-ray structures of telomeric quadruplexes reported in **Figure 43**, apart from the PDB structure 3CDM, were obtained using the sequence d[AGGG(TTAGGG)₃] which possess only a 5'-flanking adenine. The PDB structure 3CDM contains a 5'-TA flanking sequence. On the other side, all the reported NMR structures were obtained with sequences (telomeric and non-telomeric) containing more 5'- and 3'-flanking nucleotides able to form capping structures. Inspection of the NMR structures clearly shows that the 5'- and 3'-flanking segments play a major role in forming pockets able to accommodate the ligand. In particular, the complex of the *c-MYC* DNA with the quindoline derivative e) (PDB: 2L7V) is featured by an extensive rearrangement of the flanking nucleotides, that stack on top of the ligand and wrap it.

Petraccone and co-workers suggested that targeting human telomeric quadruplex multimers could be a new route for drug design.^[396] As discussed in Chapter 2.4, the 3'-end of the telomeres could form *in vivo* higher order architectures composed of more G-quadruplex units. Therefore, it would be sensible to test the ligand ability to bind not only to single G-quadruplex units, but also to G-quadruplex multimers. In fact, the binding ability to G-quadruplex multimers might be different and not predictable on the basis of the interaction data reported for the single units.

Remarks on the drug-likeness of G-quadruplex ligands

Even if characterized by excellent quadruplex over duplex selectivity and high binding affinity, most of the G-quadruplex ligands presented above do not satisfy the four Lipinski criteria for orally-administered drugs (molecular weight < 500 Daltons, octanol-water partition coefficient < 5, number of hydrogen-bond donors < 5 and number of hydrogen-bond acceptors < 5) used in medicinal chemistry to predict the drug-likeness of a candidate lead compound. If a compound has physico-chemical properties that fall in these ranges, it is likely that its adsorption, distribution, metabolism and excretion (ADME) properties will be acceptable.

The presence of large aromatic surfaces and multiple cationic charges make the pharmacokinetic properties of a classical G-quadruplex ligand rather poor.^[397]

Nevertheless, pre-clinical studies reported anti-cancer activity *in vivo* in xenograft cancer models for several G-quadruplex ligands (e.g., RHPS4, BRACO-19, TMPyP4, telomestatin, naphthalene diimide derivatives) therefore their cellular uptake must be efficient. However, it is not yet completely clear why G-quadruplex ligands show selectivity towards tumor cells lines: it may be due to differences in cell permeability or differences in composition of the shelterin complex between the normal and the cancer cells.

The potent anti-cancer activity of these small molecules in xenograft models suggests that it is worth to invest more effort in the development of G-quadruplex targeting compounds, even if they might be not optimal for oral administration.^[264]

The only G-quadruplex ligand that has entered the Clinical trials is quarfloxin (CX-3543) (see also the previous paragraphs and Chapter 4). It is now in phase II for patients with low to intermediate stage neuroendocrine carcinoma. Furthermore, several G-quadruplex ligands that did not proceed to clinical trials show synergistic effects with established anti-cancer drugs (e.g., a BRACO-19 derivative and cis-platinum) to inhibit the growth of cancer cells *in vivo*.^[397]

6.1 Research article: Spectroscopic, molecular modelling, and NMR-spectroscopic investigation of the binding mode of the natural alkaloids berberine and sanguinarine to human telomeric G-quadruplex DNA

Bessi I., Bazzicalupi C., Richter C., Jonker H. R. A., Saxena K., Sissi C., Chioccioli M., Bianco S., Bilia A. R., Schwalbe H., Gratteri P.

ACS Chemical Biology, 2012, 15; 7(6): 1109-19.

In this article, the binding ability of berberine and sanguinarine to human telomeric G-quadruplexes is discussed. The DNA-ligand interaction was studied using circular dichroism (CD), fluorescence, NMR spectroscopy and molecular modeling approaches. Our results uncover a high stoichiometry of ligand binding as well as ligand self-association induced by the interaction with G-quadruplex DNA.

NMR spectroscopy and molecular modelling were employed to gain insight into the binding modes of the two natural alkaloids with human telomeric DNA.

The NMR data were collected and analysed by the author of this thesis in the group of Prof. Schwalbe. Dr. Jonker performed the NMR-restrained docking of the ligand-quadruplex complexes. The fluorescence measurements and the CD measurements were performed in the group of Prof. Sissi (University of Padua, Italy) while the unrestrained docking was carried on in the group of Prof. Gratteri (University of Florence, Italy). The author of this thesis contributed to all the aspects of writing the manuscript.

No structural data on the complex between telomeric quadruplex and berberine or sanguinarine was available at the time this paper was written. A paper presenting the crystal structure of a human telomeric G-quadruplex with berberine was published in 2013 by Bazzicalupi *et al.* (PDB: 3R6R, see also Chapter 6).

Spectroscopic, Molecular Modeling, and NMR-Spectroscopic Investigation of the Binding Mode of the Natural Alkaloids Berberine and Sanguinarine to Human Telomeric G-Quadruplex DNA

Irene Bessi,^{†,‡,§} Carla Bazzicalupi,^{||} Christian Richter,[§] Hendrik R. A. Jonker,[§] Krishna Saxena,[§] Claudia Sissi,[⊥] Matteo Chioccioli,^{†,‡} Sara Bianco,[⊥] Anna Rita Bilia,[†] Harald Schwalbe,^{*,§} and Paola Gratteri^{*,†,‡}

[†]Department of Pharmaceutical Sciences and [‡]Laboratory of Molecular Modeling Cheminformatics & QSAR, Department of Pharmaceutical Sciences, University of Firenze, Via Ugo Schiff 6, I-50019 Sesto Fiorentino, Firenze, Italy

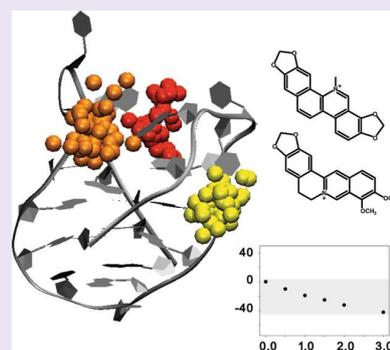
[§]Center for Biomolecular Magnetic Resonance, Institute of Organic Chemistry and Chemical Biology, Johann Wolfgang Goethe-University Frankfurt, Max-von-Laue-Strasse 7, 60438 Frankfurt/Main, Germany

^{||}Department of Chemistry "Ugo Schiff", University of Firenze, Via della Lastruccia 3-13, I-50019 Sesto Fiorentino, Firenze, Italy

[⊥]Department of Pharmaceutical Sciences, University of Padova, Via Marzolo 5, 35100 Padova, Italy

Supporting Information

ABSTRACT: G-quadruplex structures can be formed at the single-stranded overhang of telomeric DNA, and ligands able to stabilize this structure have recently been identified as potential anticancer drugs. Among the potential G-quadruplex binders, we have studied the binding ability of berberine and sanguinarine, two members of the alkaloid family, an important class of natural products long known for medicinal purpose. Our spectroscopic (CD, NMR, and fluorescence) studies and molecular modeling approaches revealed binding modes at ligand–complex stoichiometries >1:1 and ligand self-association induced by DNA for the interactions of the natural alkaloids berberine and sanguinarine with the human telomeric G-quadruplex DNA.



Telomeric DNA and its quadruplex structures are exceptional interesting targets for pharmaceutical intervention.^{1–4} Telomeres represent non-coding DNAs located at the end of eukaryotic chromosomes. In human and vertebrate cells they are composed of TTAGGG repeats. The 3'-ends of telomeres are single-stranded and in this form represent the substrate of telomerase, a reverse transcriptase that maintains telomere length homeostasis. Folding of single-stranded telomeric DNA into a G-quadruplex structure prevents hybridization with the telomerase RNA template. Inhibition of telomerase and the related telomere shortening is suggested as a therapeutic selective anticancer strategy. Indeed, telomerase is overexpressed in 80–85% of cancer cells and primary tumors but not in normal somatic cells.⁵ On the basis of this evidence, low molecular weight ligands that bind and induce telomeric G-quadruplexes are thus considered potential anticancer drugs, due to their ability to inhibit the activity of the telomerase enzyme resulting from a stabilization of the not properly folded DNA substrate. A number of synthetic compounds have been shown to either stabilize G-quadruplex structures or promote their formation.^{5,6} In addition, natural compounds are an important resource for the development of novel drugs with antitumor activity (half of all anticancer drugs

internationally approved between 1940 and 2006 are of natural origin). Several features of natural compounds make them very attractive, such as the considerable variability in their chemical structure, their wide availability from natural sources, the low cost of their extraction processes, and most importantly, their general low toxicity.⁷

Up to now, a number of studies have investigated the interaction of G-quadruplexes with small natural compounds.^{8–13} Those studies have mainly focused on the planar isoquinoline alkaloid berberine (Scheme 1a) and its derivatives. Additionally, due to its structural similarity to berberine, the benzophenanthridine alkaloid sanguinarine (Scheme 1b) has also been investigated as a potential ligand that is able to promote G-quadruplex folding from human telomeric sequences.^{11,14,15}

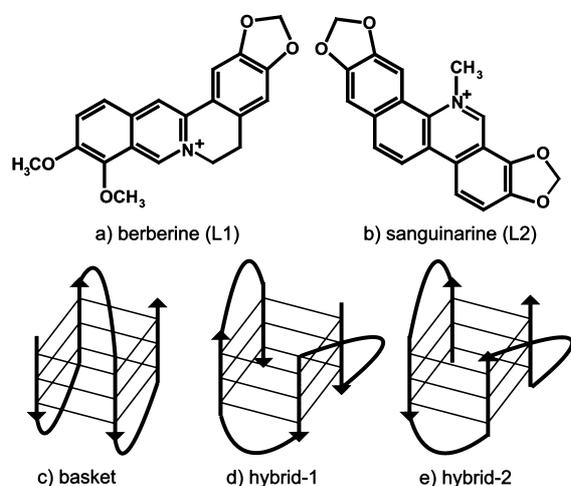
These compounds with a π -delocalized system and a partial positive charge were found to recognize G-quadruplex structures confirming that they can represent lead structures for further optimization.

Received: July 22, 2011

Accepted: April 9, 2012

Published: April 9, 2012

Scheme 1. Chemical Drawing of (a) Berberine (L1) and (b) Sanguinarine (L2) and Folding Topologies of DNA Telomeric G-Quadruplexes: (c) Basket, (d) Hybrid-1, and (e) Hybrid-2



However, to date, structural data are rare, impairing a comprehensive description of the DNA binding mode of these alkaloids. This lack of structural data is likely due to the high structural polymorphism of the human telomeric G-quadruplex DNA in solution. Short human telomeric DNA sequences, containing up to four TTAGGG repeats, are often used as models for structural investigations of the telomere, but very little is known about the biologically relevant fold of long human telomeric DNA *in vivo*.^{16,17} Indeed, the G-quadruplex fold depends on several factors, such as flanking nucleotides and nature of stabilizing cation.^{18,19} In a Na⁺-containing solution, the human telomeric DNA sequence AGGG-[TTAGGG]₃ (Tel22) assumes a unique structure, known as basket-type (Scheme 1c),²⁰ whereas in K⁺-containing solution, a mixture of different conformations can coexist.¹⁸ On the other hand, NMR spectroscopy showed that in the presence of K⁺, the AAAGGG[TTAGGG]₃AA (Tel26) sequence and the TTAGGG[TTAGGG]₃TT (wtTel26) sequence, derived from Tel22 by addition of flanking nucleotides, are folded in unique stable arrangements, the hybrid-1 and hybrid-2 topology (Scheme 1d and e), respectively.^{21,22}

In the present work, we report a comparison of the binding mode of berberine and sanguinarine to human telomeric G-quadruplex DNA adopting either basket- or hybrid-type fold. The interaction of the two alkaloids with DNA has been investigated by a variety of spectroscopic techniques (fluorescence, CD, and NMR) and molecular modeling approaches. Our investigations show that the natural alkaloids berberine and sanguinarine interact with basket- and hybrid-type human telomeric G-quadruplex DNA with a high stoichiometry as a consequence of specific binding modes and additional ligand self-association induced by DNA binding.

RESULTS

Investigation of the DNA Binding Modes by FRET Melting Studies and CD Binding Studies. The natural ligands L1 or L2 (Scheme 1) have been reported to interact with DNA,^{10,11} and to clarify whether they can discriminate between distinct DNA folds, we performed melting experiments detected by fluorescence emission changes in a FRET

dye pair on a G-quadruplex forming a sequence resembling the human telomeric sequence (4GGG) as well as on double-stranded DNA (dsDNA) sequences of comparable length. A schematic that shows the principle of the FRET/fluorescence quenching experiment is reported in Supplementary Figure 1. Under the applied experimental conditions (pH 7.4, in the presence of either Na⁺ or K⁺), the ligands interacted with the G-quadruplex DNA, as indicated by the observed increase of the DNA melting temperature. Binding of L2 induced a larger DNA melting thermal shift than L1. Ligand L2 exhibits a pH-dependent equilibrium between a charged iminium form and a neutral alkanolamine form (with a pK_a of 7.4). Since only the iminium form is known to bind strongly to B-form DNA, we evaluated whether such an equilibrium can affect the DNA recognition by L2 in conditions that are comparable to the physiological one.^{23,24} Interestingly, a pH change from 7.0 up to 8.0 did not alter the observed stabilization profile (Supplementary Figure 2).

Both ligands induced only a very small increase in the thermal stability of random dsDNA, when compared to 4GGG (Figures 1 a and b). In particular, for the human telomeric sequence, a higher ΔT_m was observed for the K⁺-stabilized G-quadruplex conformations. The melting data showed that L2 induces larger effects on all tested sequences. On the contrary, addition of unlabeled double-stranded DNA to the ligand-bound form of 4GGG displaces preferentially L2 from the G-quadruplex (Supplementary Table 1). Thus, L1 is the most selective binder for the 4GGG telomeric sequence in K⁺-containing buffer.

We additionally monitored the DNA–ligand interaction by circular dichroism (CD) spectroscopy (Figure 1 c–f). Comparable changes in the CD spectra were observed upon addition of L1 or L2 to the human telomeric DNA sequence (Tel22) in K⁺-containing buffer, suggesting a common binding mode to this hybrid-type G-quadruplex fold.²⁵ It is noteworthy that in the presence of the telomeric basket-type fold (Na⁺-containing buffer), L2 but not L1 caused relevant modifications of the DNA CD spectrum. We cannot exclude that this could be the result of a different binding mode of the two alkaloids on this G-quadruplex structure; however, merged with the melting experiments, it suggests a modest interaction of L1 with the antiparallel basket-type telomeric G-quadruplex.

Equilibrium Binding Studies by Fluorescence. To quantify the binding affinity of L2 toward DNA we performed fluorescence titrations monitoring the fluorescence of the ligand. Accordingly, we monitored L2 binding to the telomeric G-quadruplex DNA (Tel22) folded in either Na⁺- or K⁺-containing buffer (Supplementary Figure 3) as well as to calf thymus DNA (ctDNA). Data relative to the interaction of L1 with Tel22 in K⁺-containing buffer were previously reported,^{9,11,26} while in Na⁺-containing buffer, the lower binding affinity of L1 to this basket-type fold does not allow a reliable quantitative analysis. Under all conditions, addition of DNA caused quenching of the drug fluorescence signal, which allowed us to determine the binding affinity and the stoichiometry of the complex (Table 1).

Although a preference for G-quadruplexes over ctDNA can be inferred, no large modulation of the binding constant of L2 for different DNA targets was observed. This finding is in agreement with the CD and fluorescence melting data, further confirming the poor selectivity of this compound for a defined DNA arrangement. It is noteworthy that under all of the tested

Fluorescence emission quenching assays

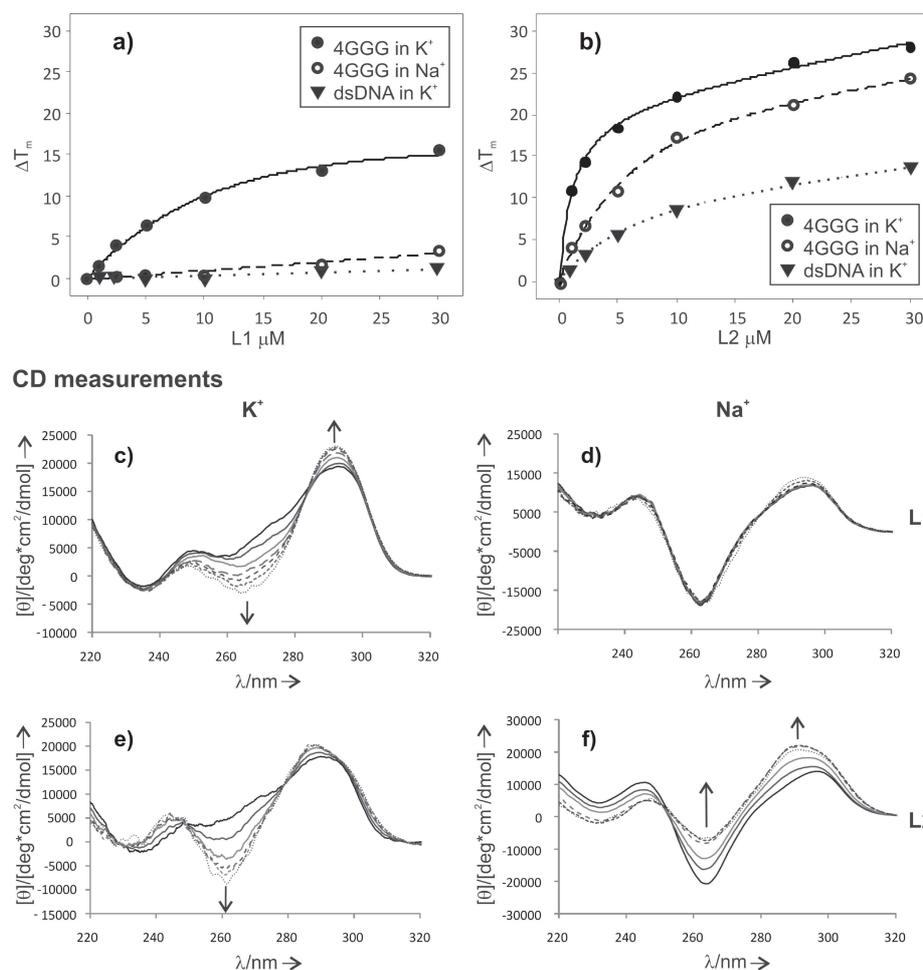


Figure 1. (a, b) Variations of DNA (0.25 μM) melting temperature by increasing concentrations of L1 (a) or L2 (b) in 10 mM LiOH, pH 7.4 with H_3PO_4 , 50 mM KCl or NaCl, evaluated by fluorescence melting experiments. Error ± 0.4 °C. (●) 4GGG in K^+ -containing buffer; (○) 4GGG in Na^+ -containing buffer; (▼) dsDNA in K^+ -containing buffer. (c–f) Circular dichroism spectra of Tel22 (4 μM) recorded in the presence of increasing L1 (c, d) or L2 (e, f) concentrations (0–50 μM) recorded in 10 mM Tris-HCl, pH 7.5 with 50 mM KCl (c, e) or 50 mM NaCl (d, f).

Table 1. Thermodynamic Parameters from Fluorometric Titrations Performed in 10 mM Tris, pH 7.5 with Added 50 mM NaCl or KCl

target DNA	buffer	$K_b \times 10^6$ (M^{-1})	stoichiometry L2/Tel22	n (bases) ^a
Tel22	K^+	9.14 ± 0.85	2.0 ± 0.1	1.7 ± 0.2
	Na^+	2.45 ± 0.21	6.0 ± 0.4	
ctDNA	K^+	1.76 ± 0.15		

^aThe number of DNA residues involved in the interaction with a single ligand molecule.

conditions a binding stoichiometry higher than 1:1 was detected.

Insights into the Structure of the Complex via NMR Spectroscopy and Molecular Modeling. The NMR experiments were planned in order to characterize the interaction of berberine and sanguinarine with antiparallel basket-type as well as hybrid-type folds, which were found to be the prevalent conformations of intramolecular telomeric G-quadruplexes in physiologically relevant conditions.¹⁷

The experimental conditions and the DNA sequences were chosen to achieve optimal NMR results: the Tel22 sequence was used in Na^+ -containing buffer while the Tel26 and wtTel26 sequences were used in K^+ -containing buffer. The human telomeric Tel22 sequence AGGG[TTAGGG]₃ in Na^+ -containing solution forms a unique basket-type structure,²⁰ while in the presence of K^+ , an equilibrium is observed between more hybrid-type structures that are energetically comparable, as indicated by 1D ¹H NMR spectra (Supplementary Figures 4 and 5). Simple modifications of the Tel22 sequence lead to a single prevalent structure in K^+ -containing solution: the Tel26 sequence AAAGGG[TTAGGG]₃AA and the wtTel26 sequence TTAGGG[TTAGGG]₃TT are folded into the so-called hybrid-1 and hybrid-2 structures, respectively.^{21,22}

The ¹H resonances assignment of the ligands (Supplementary Figures 6–9 and Tables 2–3) is in accordance with data reported in the literature.^{27,28} The proton resonance assignment for the different DNA folds was performed on the basis of the assignment as previously reported in literature.^{20–22} The ¹H NMR spectra of Tel22 with assignment of imino region and a ³¹P NMR proton decoupled spectrum of Tel22 with partial

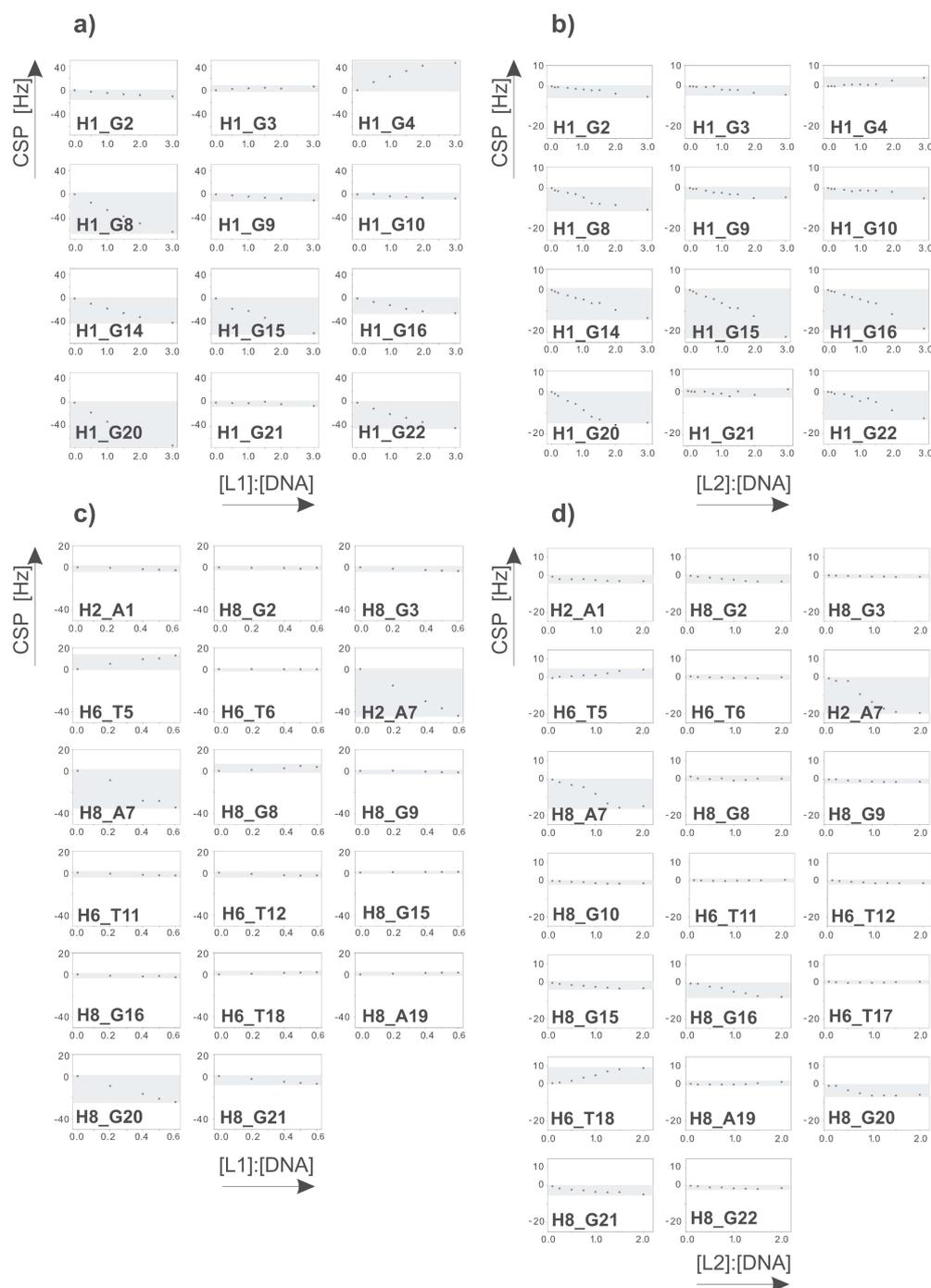


Figure 2. Chemical shift perturbation (CSP) as function of [ligand]:[DNA] molar ratio. (a) CSP of imino protons versus [L1]:[DNA]. (b) CSP of imino protons versus [L2]:[DNA]. (c) CSP of aromatic protons versus [L1]:[DNA]. (d) CSP of aromatic protons versus [L2]:[DNA]. (a, c) Titration of L1 up to a [L1]:[DNA] molar ratio equal to 0.6 because the most significant CSP were observed in this range of [L1]:[DNA].

assignment of nucleotides involved in the interaction are included in Supplementary Figures 10–11.

To gain an insight into the binding modes between the ligands and the DNA, NMR titrations of the selected different G-quadruplexes with L1 and L2 were performed. Due to signal overlap, the chemical shift perturbations (CSPs) of Tel26 and wtTel26 could not accurately be monitored (NMR titration data relative to the interaction of berberine with Tel26 and wtTel26 is presented in Supplementary Figure 12), but

nevertheless, the broadening of some external G-quartet imino signals can be clearly observed. Therefore, our NMR studies mainly focused on the investigation of the interaction between the ligands and the Tel22 in basket-type conformation.

The CSPs as function of the [ligand]:[DNA] ratio for imino and aromatic proton resonances are displayed in Figure 2.

Imino proton signals that showed the most intense CSP upon addition of L1 are G4, G8, G14–G15–G16, G20, and G22 (Figure 2a). Due to signal overlap in the aromatic region of

proton spectrum, not all of the aromatic proton chemical shifts could be followed during the titration. G20 (H8) and T5 (H6) are significantly perturbed as well (Figure 2c).

For L2, the most perturbed imino protons belong to the groove defined by the adjacent G14-G15-G16 and G20-G21-G22 stretches, suggesting that this side of Tel22 G-quadruplex should be the one mainly involved in the interaction with L2 (Figure 2b). Among the monitored aromatic protons, A7 (H2, H8) displays a continuous significant resonance shift, up to a saturation level around a molar ratio $[L2]:[Tel22] = 1.5$; G16 (H8), T18 (H6), and G20 (H8) show relevant induced chemical shifts as well (Figure 2d). For both ligands, an intense broadening of A19 H2 signal can be observed during the titration.

It is noteworthy that the range of chemical shift (δ) variation due to addition of L1 is much larger than that observed upon addition of L2. For example, imino proton of G20 at a ratio ligand:DNA = 3:1 gives rise to a $\Delta\delta \approx 15$ –20 Hz in presence of L2, whereas in presence of L1 the same proton, under the same conditions, is characterized by a $\Delta\delta \approx 75$ Hz. The extent of the CSP may stem from different effects, such as the different degree of aromaticity of the ligands and thus differences in the induced aromatic ring shifts.

Figure 3 summarizes the chemical shift perturbation data obtained by NMR titrations. On the basis of the ^1H CSPs

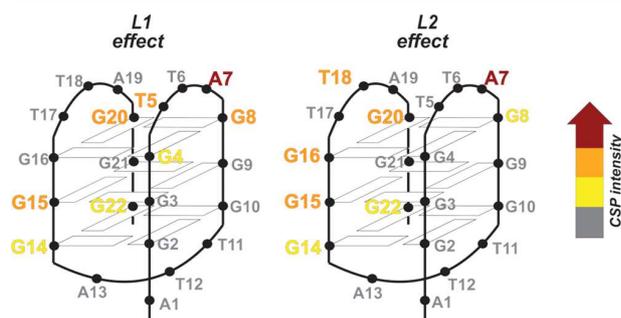


Figure 3. Ligands effect on Tel22 imino and aromatic proton chemical shift derived from NMR titrations in Na^+ -containing solution.

mapping, the topology of the binding sites is similar for both ligands.

In order to obtain further information concerning the location of the binding sites, proton-decoupled 1D ^{31}P NMR spectra of Tel22 and the Tel22–ligand complexes were recorded. ^{31}P CSPs between the complex and DNA are displayed in Figure 4a. Among the residues that could unambiguously be monitored, phosphates 5 and 17 seem to be most perturbed by L1, whereas in the Tel22:L2 complex, residues 9 and 20 revealed the largest ligand-induced chemical shift perturbation, in accordance with binding regions as indicated by the ^1H CSPs.

$^1\text{H}, ^1\text{H}$ -NOESY and $^1\text{H}, ^{13}\text{C}$ -HSQC spectra were recorded on the Tel22:L2 complex since its higher stability (compared to the Tel22:L1) allowed for longer measurements needed to obtain high quality 2D NMR spectra. $^1\text{H}, ^1\text{H}$ -NOESY spectra of the Tel22:L2 complex (Figure 4b) were recorded using different mixing times (100, 300, and 500 ms). The region connecting base aromatic protons with sugar H1' protons shows that the most shifting peaks involve residues A7, T18, and G20, as already evidenced by NMR titration experiments.

Under the applied experimental conditions, no NOE cross peaks between ligand and DNA were observed.

The overlay of the natural abundance 2D $^1\text{H}, ^{13}\text{C}$ -HSQC spectra of free Tel22 with Tel22:L2 complex (Figure 4c) revealed that the largest chemical shift involves H8–C8 of adenine A1. Furthermore, no signal of the free ligand is detected under the applied experimental conditions. Thus, in order to determine chemical shift perturbations of the ligand and which ligand atoms are mainly involved in the interaction with the G-quadruplex, a titration of L2 upon addition of DNA was performed (Figure 4d). Ligand signals become broader and disappear already at $[L2]:[Tel22]$ ratio of around 30. Precipitation as a potential source for the disappearance of ligand signals was not observed, even after weeks.

Molecular Modeling. Two different modeling approaches were used in this work, a NMR-driven and an unrestrained docking approach. Modeling of the adducts between the ligands and the telomeric DNA in basket-type and hybrid-type conformation was performed using unrestrained docking and validated by NMR-driven docking of L1 and L2 with basket-type DNA. This strategy was chosen because NMR titrations provided unambiguous information only for the quadruplex in basket-type conformation.

Unrestrained Docking: Basket Structure. For the basket-type DNA, the unrestrained docking procedure pointed out three and four interaction sites for L1 and L2, respectively (Figure 5a and b). It is worth noting that two of these binding sites coincide with those observed by NMR-driven docking (orange and yellow color codes in Figure 5). In addition, the A7 adenine residue moves from the original position occupied in the unbound DNA structure and associates *via* H-bond with A19, forming an additional platform available for $\text{CH}\cdots\pi$ or $\pi\cdots\pi$ interactions in the case of L1 and L2, respectively (Supplementary Figure 13). The area of the basket-type diagonal loop hosts another interaction site: L1 mainly interacts with the A13 and additionally with G22 *via* $\text{CH}\cdots\pi$, while L2 is involved in $\text{CH}\cdots\pi$ interaction with A1 (blue molecules in Figure 5a and b). The results obtained for L2 are in agreement with the 2D $^1\text{H}, ^{13}\text{C}$ -HSQC spectral data reported in Figure 4c. In the case of L2, which features a higher stoichiometry ratio than L1, a further localized binding site involves the G8-G9-G19 and G2-G3-G4 stretches (green molecule in Figure 5b).

Unrestrained Docking: Hybrid Structure. An unrestrained docking procedure was used to explore the behavior of both ligands with hybrid-type DNA structures. The lowest energy predicted structures for the ligand–DNA hybrid folding adducts are shown in Figure 5c and d. In the case of L1, the blue ligand is stacked on the 5'-end G-tetrad, interacting *via* $\pi\cdots\pi$ stacking with the G10 and *via* $\text{CH}\cdots\pi$ with G4 residues (Supplementary Figure 14). This binding mode is in agreement with the modeling simulation previously reported by Arora *et al.*⁹ for the L1:G-quadruplex adduct in K^+ -containing solution showing the L1 molecule stacked on an external G-quartet. Furthermore, L1 can additionally achieve good π -stacking with the G6 residues from the 3'-end G-tetrad and with A26 (orange molecule in Figure 5c; also Supplementary Figure 14). Moreover, a $\text{CH}\cdots\pi$ interaction is observed with the methyl group of the T7 residue. The results obtained in the case of the complex formed between L2 and the hybrid-type structure are shown in Figure 5d. As observed for the model of L1:DNA hybrid conformation, two poses were found characterized by about 15 kcal/mol energy gap, and in one of them L2 interacts with the 5'-end G-tetrad of the hybrid G-quadruplex (orange

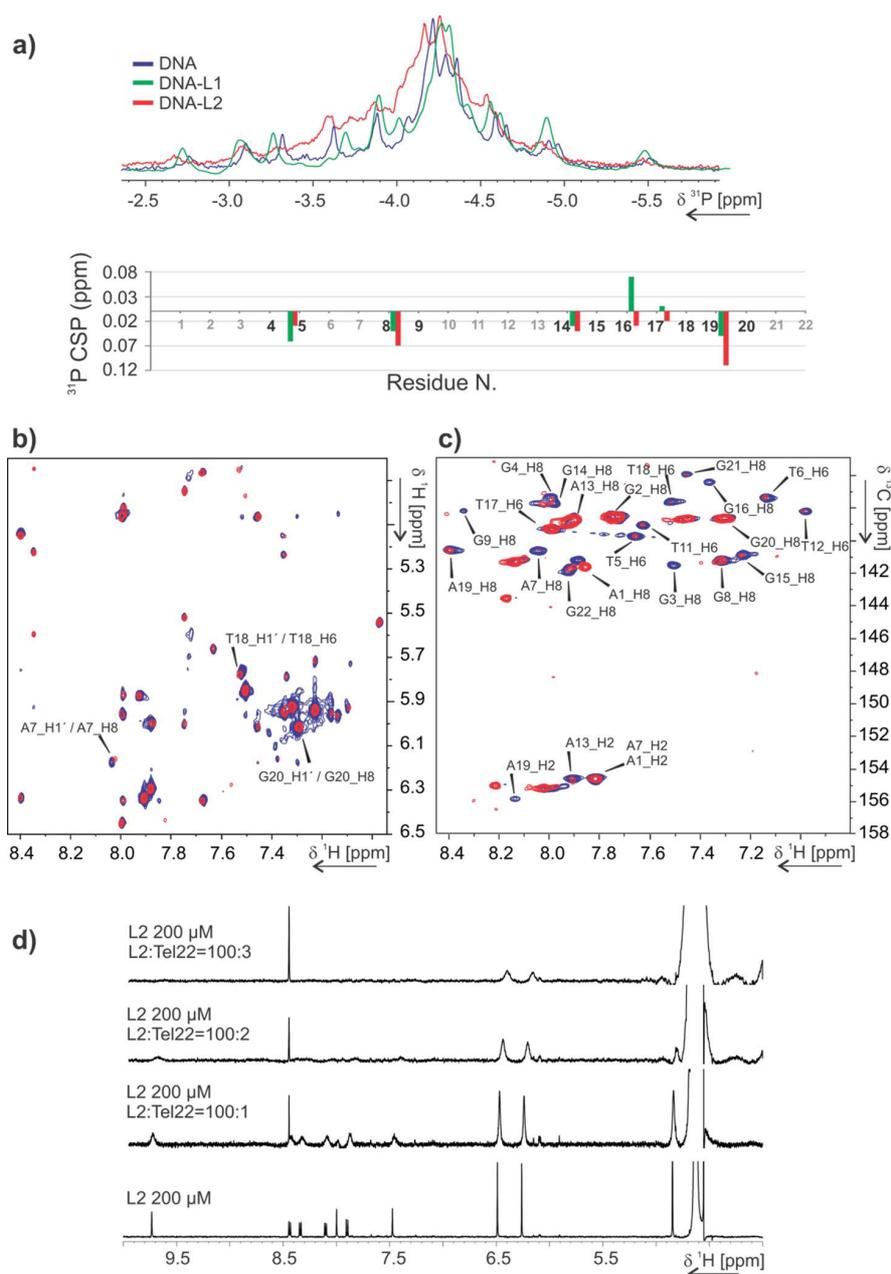


Figure 4. (a) Top: 1D ^1H -decoupled ^{31}P NMR spectrum of Tel22:L1 = 1:1.5 complex 200 μM DNA in Na^+ -containing buffer, 90% $\text{H}_2\text{O}/10\%$ D_2O , 25 $^\circ\text{C}$, 300 MHz (green) and 1D ^1H -decoupled ^{31}P NMR spectrum of Tel22:L2 = 1:2 complex 200 μM DNA in Na^+ -containing buffer, 90% $\text{H}_2\text{O}/10\%$ D_2O , 25 $^\circ\text{C}$, 300 MHz (red) overlaid with 1D ^1H -decoupled ^{31}P NMR spectrum of Tel22 370 μM in Na^+ -containing buffer, 100% D_2O , 25 $^\circ\text{C}$, 300 MHz (blue). Bottom: ^{31}P CSPs due to the ligands are plotted as a function of the residue number N. (b) Overlay of $^1\text{H},^1\text{H}$ -NOESY of the complex (red, Tel22:L2 = 1:2, Tel22 1 mM, 25 $^\circ\text{C}$, 950 MHz) with the DNA alone (blue, Tel22 370 μM , 25 $^\circ\text{C}$, 600 MHz) in Na^+ -containing buffer, 100% D_2O . (c) Overlay of the $^1\text{H},^{13}\text{C}$ -HSQC of the complex (red, Tel22:L2 = 1:2, Tel22 300 μM , 25 $^\circ\text{C}$, 700 MHz) with the DNA alone (blue, Tel22 1 mM, 25 $^\circ\text{C}$, 700 MHz) in Na^+ -containing buffer, 10% $\text{D}_2\text{O}/90\%$ H_2O . (d) Titration of 200 μM L2 with Tel22 in Na^+ -containing buffer 100% D_2O , 37 $^\circ\text{C}$, 700 MHz.

ligand in Figure 5d; also Supplementary Figure 15) and in the other inserts in one lateral groove (yellow ligand in Figure 5d). It is to be noted that the most stable calculated pose for this adduct is the one placed in the groove (yellow molecule, Figure 5d), whereas the other, most likely due to the steric hindrance of the $\text{N}-\text{CH}_3$ group, mainly interacts with only one guanine residue of the G-tetrad (orange molecule in Figure 5d).

NMR-Driven Docking. Restrained docking based on NMR data of ligand molecules onto the basket-type G-quadruplex DNA structure clearly indicates the presence of different possible binding sites, in accordance with the hypothesis of multiple binding sites having similar energy derived by spectroscopic titrations. Since NMR titrations did not provide atom specific ligand restraints (since the signals of the ligand are directly lost at the beginning of the titration) and

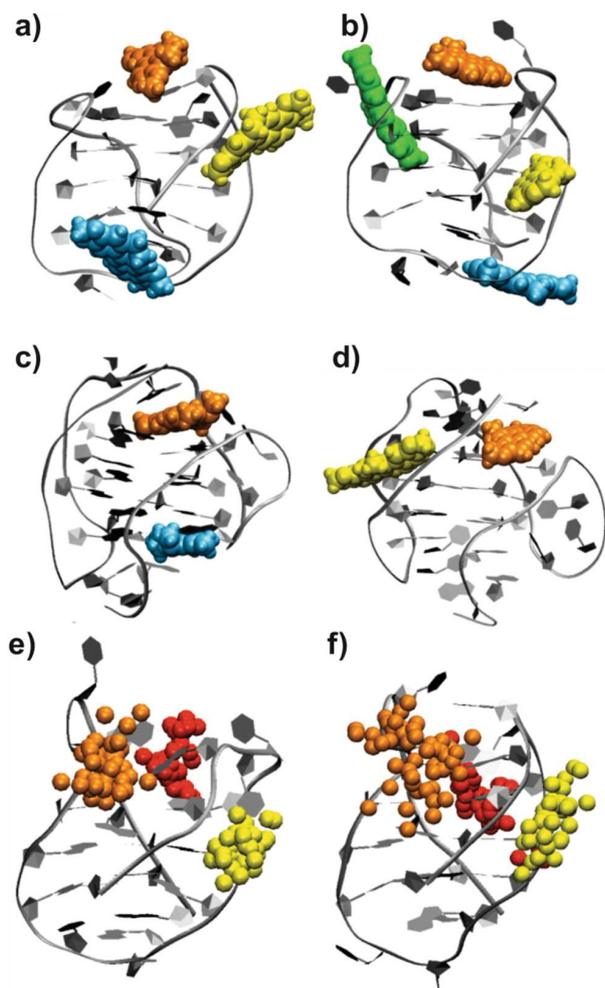


Figure 5. Docking results. Unrestrained-docking poses within 15 kcal/mol from the lowest in energy, obtained for the adducts between L1 and basket-type G-quadruplex DNA (a) and L2 and basket-type G-quadruplex DNA (b). Unrestrained-docking poses within 15 kcal/mol from the lowest in energy, obtained for the adducts between L1 and hybrid-1 type DNA (c) and L2 and hybrid-1 type DNA (d). Overlay of the 50 low-energy representative NMR-driven docking models obtained for the basket-type G-quadruplex DNA with three ligand molecules of the Tel22:L1 complex (e) and the Tel22:L2 complex (f). The ligands are represented as their centroid.

consequently the docking calculations could not produce structures with a unique ligand orientation, only the centroid of each ligand molecule is displayed in Figure 5e and f. The three observed binding sites are similar for both ligands: one binding site is located close to residue A7 (orange centroids in Figure 5e and f), another interaction site is located in the groove defined by G14-G15-G16 and G20-G21-G22 stretches (yellow centroids in Figure 5e and f), and the third binding site (red centroids in Figure 5e and f) is situated either close to the handles of the basket (residues G16-T17 and G4-G3) or toward the G2-G3-G4/G14-G15-G16 groove residues, respectively, for L1 and L2. Two binding sites predicted by unrestrained docking are observed also by NMR-driven docking (orange and yellow color codes in Figure 5). Nevertheless, based on our spectroscopic studies, a higher binding stoichiometry in the formation of these adducts has to

be considered, which may not clearly be monitored by the NMR titrations and which becomes far too ambiguous for NMR-based multiple-ligand docking calculations. The ambiguity raises dramatically as either more binding sites or multiple stacking of ligand molecules in one binding site would be possible.

DISCUSSION

From spectroscopic titrations there is evidence for higher than 1:1 stoichiometry: up to 6 ligand molecules (n) bind to one DNA molecule, depending on the experimental conditions. A different behavior for the two investigated ligands was recently reported that might likely be due to the use of different buffer conditions.¹¹ Fluorescence data revealed a higher stoichiometry for the basket-type than for the hybrid-type folds (determined in Na⁺- or K⁺-containing buffer, respectively). We propose that the number of bound molecules determined by fluorescence titration arises from the sum of different energetically similar binding events. Under almost all tested conditions, the complex stoichiometry did not correspond to a simple interaction mode *via* π -stacking with the external G-tetrad, which would most likely involve two ligand molecules per intramolecular telomeric quadruplex. This is remarkably pronounced in the presence of Na⁺ where experimental data can be satisfactorily fitted only taking into account a large number of ligands per G-quadruplex unit. A possible explanation might be the formation of unselective interactions corresponding to an external drug binding mode promoted by charge interaction. However, the NMR titration results obtained in Na⁺-containing buffer indicate a number of specific binding sites. The overall shape of the 1D ³¹P NMR spectrum is maintained after adding ligand, and only a limited set of residues is characterized by CSP in the 1D ¹H spectrum, while the others remain clearly unperturbed.

In order to reconcile the experimental findings, we propose the following hypothesis. The binding stoichiometry is the result of two different concurring interaction processes: the first involves a number of molecules (n) at least equal to the number of binding sites (m) that specifically interact with the G-quadruplex structure; the latter is the self-association of ligand induced by DNA quadruplex binding and accounts for the remaining interacting molecules ($n - m$).

Self-association of ligand induced by DNA binding is supported by the fact that no signals of free/bound ligand in complex were detected by NMR. In addition, during NMR titration of L2 with Tel22 quadruplex, broadening of ligand signals already occurs at a [ligand]:[DNA] ratio of 100:1 (Figure 4d).

It is noteworthy that a high binding stoichiometry to G-quadruplex structures was previously also observed for porphyrines and was attributed to self-association of the ligand, after binding to G-quadruplex DNA.²⁹ Most of the small molecules that bind to DNA are planar aromatic compounds of considerable hydrophobic character. Compounds with these properties turn out to be good intercalators and/or minor groove binders but also favor self-association in aqueous solution.³⁰ The extent of self-association is influenced in particular by the ligand and DNA concentration and by several other factors, such as buffer ionic strength and pH. Therefore this phenomenon cannot be ignored especially when the experimental technique (such as NMR) requires a relatively high concentration.

The binding stoichiometry analysis revealed that the DNA-promoted self-association of L2 is much lower in K⁺-containing

buffer (hybrid-type G-quadruplex) than in Na⁺-containing buffer (basket-type G-quadruplex). This fact could be explained considering the ligand surface accessibility in the bound state: in the basket-type at least one ligand is observed in an external binding site (orange molecules nearby A7, Figure 5a and 5b), whereas in the case of the complexes with hybrid-type, the ligand surface is less available for further stacking.

The little structural data available frequently refer to a human telomeric quadruplex in propeller-type conformation and present the ligand stacking directly onto the external G-quartets.^{31–33} According to our NMR data, we do not find evidence that L1 and L2 stack directly onto the external G-quartets of the basket-type G-quadruplex (Figure 5a and b).

The external G-tetrads in the Tel22 basket-type fold are much more hidden when compared to propeller-type conformation and therefore not so easily accessible for the ligand molecules. Our model suggests that, for the basket-type G-quadruplex fold, the loops and grooves play an important role in ligand recognition.

For both ligands, a higher increase of the melting temperature was observed for the hybrid-type (in K⁺-containing buffer) than for the basket-type (in Na⁺-containing buffer) fold. Beside the direct metal ion effect on DNA stability, this behavior could reflect a preferred interaction. Ligand L2 appears to be more efficient and L1 more specific in recognizing the different G-quadruplex structure folds. On the basis of our structural results, we can hypothesize that the larger effect on the quadruplex thermal stability observed in K⁺-containing solution might be due to the different nature of the binding interactions between ligand and different quadruplex folds. Both L1 and L2 are able to externally stack onto the G-tetrads only when complexed with hybrid-type G-quadruplex structures, whereas for the basket-type structure they can only interact with grooves and loops, mostly *via* π -stacking with the aromatic moiety of a single base.

In summary, our detailed spectroscopic investigations revealed high stoichiometry specific binding modes and additional self-association for the interactions of the natural alkaloids berberine and sanguinarine toward basket-type and hybrid-type human telomeric G-quadruplex DNA. Binding interactions are found to be influenced by several conditions. The multivalent binding and the DNA-induced ligand self-aggregation were not previously described for these ligands and could be used in advanced pharmaceutical anticancer drug research to design analogues with improved specificity to refine the inhibition of the activity of the telomerase enzyme.

METHODS

Material. Berberine and sanguinarine, indicated as L1 and L2, respectively, were purchased from Sigma-Aldrich Co. Ltd. (USA) and used without further purification.

NMR experiments were performed using intramolecular G-quadruplex formed by d[AG₃(T₂AG₃)₃] (Tel22), d[AAAGGG-(TTAGGG)₃AA] (Tel26), and d[TTAGGG(TTAGGG)₃TT] (wtTel26), which were purchased by Eurofins MWG Operon (Ebersberg, Germany) as HPSF (High Purity Salt Free) purified oligos and further purified *via* HPLC.

DNA thermal stability was measured by using the sequence 5' Dabcyl-AG₃(T₂AG₃)₃-T-FAM 3' synthesized and purified by ATDBIO (Southampton, U.K.) (4GGG) and the double-stranded fragment (dsDNA) derived from the annealing of the 5' FAM-GTGA-TACCGACAGAAG and CTTCTGTCGGTATCTCAC- Dabcyl 3' sequences purchased from Eurogentec (Belgium).

Calf thymus DNA (ctDNA) was purchased by Sigma-Aldrich (USA) and used with no further purification.

For fluorescence melting experiments, fluorescence titrations and CD measurements, DNA solutions in the required buffer were heated at 95 °C for 5 min and left to cool at RT overnight. DNA samples used for NMR measurements were prepared using a snap cool annealing procedure in order to avoid the formation of multimers.

Fluorescence Quenching Assays. Fluorescence melting experiments were performed in a Roche LightCycler, using an excitation source at 488 nm. The changes in fluorescence emission were recorded at 520 nm. Melting experiments were performed in a total volume of 20 μ L containing 0.25 μ M DNA and variable concentrations of tested derivatives in 10 mM LiOH, pH 7.4 with H₃PO₄, and 50 mM KCl or NaCl. Recordings were taken during a heating step from 30 to 95 °C at 1 °C/min.

T_m values were determined from the first derivatives of the melting profiles using the Roche LightCycler software. Each curve was repeated at least three times, and errors were \pm 0.4 °C.

CD Measurements. Circular dichroism spectra from 200 to 350 nm were recorded at 25 °C using 1 cm path length quartz cells on a Jasco J 810 spectropolarimeter equipped with a NESLAB temperature controller.

CD spectra were recorded on samples containing 4 μ M Tel22 and increasing ligands concentrations in 10 mM Tris-HCl, 50 mM KCl or 50 mM NaCl, pH 7.5. The reported spectrum of each sample represents the average of 3 scans. Observed ellipticities were converted to mean residue ellipticity [θ] = deg \times cm² \times dmol⁻¹ (Mol. Ellip.).

Spectrofluorimetric Titration Experiments. Fluorometric titrations were performed at 25 °C in 10 mM Tris-HCl, 50 mM KCl or NaCl, pH 8, with a Perkin-Elmer LS30 fluorometer, equipped with a Haake F3-C thermostat. Binding was followed by addition of increasing amounts of DNA to a freshly prepared drug solution. To avoid large systematic inaccuracies due to experimental errors, the range of bound drug fractions utilized for calculations was 0.15–0.85. All data were analyzed according to Scatchard equation. For ctDNA, data were evaluated according to the equation of McGhee and Von Hippel for non-cooperative ligand–lattice interactions:

$$\frac{r}{m} = \frac{K_b(1 - nr)^n}{[1 - (n - 1)r]^{n-1}} \quad (1)$$

where r is the molar ratio of bound ligand to DNA, m is the free ligand concentration, K_b is the intrinsic binding constant, and n is the exclusion parameter.³⁴

NMR Measurements. All NMR samples were referenced with 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) and prepared either in 25 mM Na-phosphate buffer containing 70 mM NaCl (pH = 7.0) or 25 mM K-phosphate buffer containing 70 mM KCl (pH = 7.0). All samples were dissolved in either 90% H₂O/10% D₂O or 100% D₂O. D₂O NMR samples were prepared by repeated freeze-drying and dissolving in D₂O (99.98% D).

The final NMR samples contained 0.015–1 mM DNA or 0.2–0.8 mM ligand in NMR buffer.

NMR spectra of ligands were recorded on Bruker 400 MHz spectrometer at 25 °C. 1D ¹H NMR employed water suppression schemes using excitation sculpting with gradients,³⁵ and ¹H,¹H-ROESY³⁶ (200 ms mixing time) and ¹H,¹³C-HSQC^{37–39} experiments were performed for the resonance assignment of the NMR signals of L1 and L2.

NMR spectra of Tel22, Tel22:ligand complexes and NMR titrations were recorded at 25 °C on Bruker 600, 700, 800, 900, and 950 MHz spectrometers.

¹H NMR spectra were recorded on DNA and DNA:ligand complexes with gradient-assisted excitation sculpting for water suppression³⁵ or jump-return-Echo.⁴⁰ ¹H,¹³C-HSQC^{37–39} and 2D ¹H,¹H-NOESY spectra with different water suppression schemes (Watergate^{41,42} and excitation sculpting with gradients³⁵) and different mixing times (100, 300, and 500 ms) were recorded.

DNA was titrated at 25 °C with ligands up to an excess ratio equal to 3.0 (DNA:ligand). Titration experiments were monitored by 1D ¹H

NMR spectra. 1D ^{31}P NMR spectra were recorded at RT both on Tel22 and Tel22:ligand complex samples on a Bruker 300 MHz spectrometer.

All NMR data were collected, processed, and analyzed using the software TopSpin 2.1 (Bruker) and Sparky 3.⁴³ The NMR titration curves were made using the software SigmaPlot 11.0.

Molecular Modeling. Unrestrained Docking. The binding ability of L1 and L2 (iminium form) has been investigated toward Tel22 basket-type (PDB code: 143D),²⁰ Tel26 hybrid-1 (PDB code: 2HY9),²¹ and wtTel26 hybrid-2 (PDB code: 2JPZ),²² telomeric G-quadruplex structures. Both ligand molecules have been built by the Build module of Maestro v. 8.5.⁴⁴ Docking calculations were performed using Glide v.4.5⁴⁵ with the DNA structures kept fixed in their original conformations throughout the docking procedures. Selected poses of the ligand/target (both basket or hybrid fold topologies) complexes were submitted to molecular dynamics simulation for 10 ns in explicit solvent and the root-mean-square deviation (rmsd) of the complexes was monitored *versus* simulation time; 500 cycles of minimization in water implicit solvent treatment (GB/SA) were then applied to all the stable poses (MacroModel, OPLS2005).⁴⁶

The atomic electrostatic charges of the ligands were calculated by means of the RESP procedure,⁴⁷ that is, fitting them to an electrostatic potential calculated at the HF/6-31G* level of theory using the Gaussian09 software.⁴⁸ Then, general amber force fields (GAFF) parameters were assigned to the ligands by the *antechamber* module implemented in AMBER9 suite.⁴⁹ Each complex was immersed in a truncated octahedral box, whose edges were located 10 Å apart from the closest atom of the DNA fragments and contained about 3500 water molecules. To maintain neutrality of the system, 18 Na⁺ and 22 K⁺ counterions were added to the solvent bulk of the basket-type and of the hybrid-type G-quadruplex–water complexes, respectively. In the case of the basket-type structure, two Na⁺ ions were placed along the axis within the central core of the complex, whereas in the case of the hybrid-type structures two K⁺ ions were placed along the axis within the central core of the complex, midway between each G-tetrad. Before starting the MD simulations, a minimization of the complexes was performed by setting a convergence criterion on the gradient of 0.01 kcal mol⁻¹ Å⁻¹. Then, water shells and counterions were equilibrated for 40 ps at 27 °C, and subsequently, 10 ns of MD simulations in isothermal–isobaric ensemble were performed without any restraint on each complex. The ff03 version of the AMBER force field was used for the DNA fragments and the counterions,⁵⁰ whereas the TIP3P model⁵¹ was employed to explicitly represent water molecules. In the production runs, the ligand–DNA fragments systems were simulated under periodic boundary conditions. van der Waals and short-range electrostatic interactions were estimated within a 10 Å cutoff, whereas the long-range electrostatic interactions were assessed by using the particle mesh Ewald method,⁵² with 1 Å charge grid spacing interpolated by fourth-order B-spline and by setting the direct sum tolerance to 10⁻⁵. Bonds involving hydrogen atoms were constrained by using the SHAKE algorithm⁵³ with a relative geometric tolerance for coordinate resetting of 0.00001 Å. Berendsen's coupling algorithms⁵⁴ were used to maintain constant temperature and pressure with the same scaling factor for both solvent and solutes and with the time constant for heat bath coupling maintained at 1.5 ps. The pressure for the isothermal–isobaric ensemble was regulated by using a pressure relaxation time of 1 ps in the Berendsen's algorithm. The simulations of the solvated complexes were performed using a constant pressure of 1 atm and a constant temperature of 27 °C. A time step of 2 fs was used in the simulations, which were carried out with the AMBER9 program suite.⁴⁹

NMR-Driven Docking. Modeling of the Tel22 basket in complex with the ligands L1 or L2 was achieved using a high ambiguity driven docking approach with the program HADDOCK 2.1.^{55,56} The ambiguous interaction restraints (AIRs) were defined from NMR titration experiments for the nucleotides that exhibited significant chemical shift changes upon interaction with the ligand (see Supplementary Tables 4 and 5).

Proton signals that show strong, medium, and weak chemical shift perturbations were restrained to be within a distance of 4.0, 5.0, and 6.0 Å from the ligand(s), respectively. The docking calculations were performed using the coordinates from the NMR structure bundle of the free form DNA quadruplex (PDB code: 143D).²⁰

The topology and parameter files for the ligand were manually generated using initial suggestions from the PRODRG server.⁵⁷ Furthermore, hydrogen bond restraints have been added (for all except the ones between nucleotides A7-A19, G15-G21, G8-G20 and G20-G16) to keep the secondary structure of the DNA intact during the water refinement. Previous to the docking, arbitrary starting structures were generated from the bundle of 6 DNA structures with up to 4 ligand molecules. In the first iteration, a rigid body docking was performed (7680 structures) starting at 2000 K (500 steps) and cooling down to 500 K (500 steps). Then, semiflexible docking annealing stages were performed in which the semiflexible segments were automatically defined by HADDOCK from 1000 to 50 K (1000 steps) with flexible side-chains at the interface followed by a final annealing stage from 500 to 50 K (1000 steps) with a fully flexible interface. The final lowest energy structures were further refined in explicit water⁵⁸ using the nucleic acid forcefield with OPLS charges and Lennard-Jones nonbonded parameters.^{59–61} Initial results obtained from docking of 1 and 2 ligand molecules were used to fairly define the binding sites for docking with 3 ligand molecules. The 200 final water-refined docking results were analyzed and clustered (the 3-ligand dockings of berberine and sanguinarine clustered well within a threshold of 5.0 Å and 5.2 Å rmsd, respectively). The top-ranked ensemble, according to the average interaction energy and buried surface area, was accepted as the best representative of the complex.

■ ASSOCIATED CONTENT

📄 Supporting Information

Effect of pH on melting temperature by fluorescence quenching assays; fluorescence titration data; 1D ^1H NMR spectra of L1, L2, including proton and carbon resonance assignment; 1D ^1H NMR spectra of Tel22 in different experimental conditions; ^1H , ^1H -NOESY of the imino region of Tel22 at different temperatures; 1D ^{31}P NMR spectrum of Tel22, including ^{31}P resonance assignment; titration of Tel26 and wtTel26 monitored by 1D ^1H NMR spectra; unrestrained molecular modeling detailed results; definition of ambiguous interaction restraints (AIRs). This material is available free of charge *via* the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

✉ Corresponding Author

*E-mail: schwalbe@nmr.uni-frankfurt.de; paola.gratteri@unifi.it.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors are grateful to the European community (EU-NMR and BIO-NMR). Financial support by the Access to Research Infrastructures activity in the seventh Framework Programme of the EC (Project numbers 261863, Bio-NMR and 261572, WeNMR) for conducting the research is gratefully acknowledged. The state of Hessen (BMRZ) and the Italian Ministero dell'Università e della ricerca are acknowledged for financial support. Ente Cassa di Risparmio di Firenze, Italy is also gratefully acknowledged for a grant to M.C. H.S. is member of the DFG-funded cluster of excellence: macro-molecular complexes.

REFERENCES

- (1) Balasubramanian, S., and Neidle, S. (2009) G-quadruplex nucleic acids as therapeutic targets. *Curr. Opin. Chem. Biol.* 13, 345–353.
- (2) Ou, T. M., Lu, Y. J., Tan, J. H., Huang, Z. S., Wong, K. Y., and Gu, L. Q. (2008) G-quadruplexes: Targets in anticancer drug design. *ChemMedChem* 3, 690–713.
- (3) 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.
- (4) Lipps, H. J., and Rhodes, D. (2009) G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol.* 19, 414–422.
- (5) De Cian, A., Lacroix, L., Douarre, C., Temime-Smaali, N., Trentesaux, C., Riou, J. F., and Mergny, J. L. (2008) Targeting telomeres and telomerase. *Biochimie* 90, 131–155.
- (6) Arora, A., Kumar, N., Agarwal, T., and Maiti, S. (2010) Human telomeric G-quadruplex: targeting with small molecules (Retraction of vol 6, pg 627, 2008). *FEBS J.* 277, 1345.
- (7) Efferth, T., Li, P. C. H., Konkimalla, V. S. B., and Kaina, B. (2007) From traditional Chinese medicine to rational cancer therapy. *Trends Mol. Med.* 13, 353–361.
- (8) Franceschin, M., Rossetti, L., D'Ambrosio, A., Schirripa, S., Bianco, A., Ortaggi, G., Savino, M., Schultes, C., and Neidle, S. (2006) Natural and synthetic G-quadruplex interactive berberine derivatives. *Bioorg. Med. Chem. Lett.* 16, 1707–1711.
- (9) Arora, A., Balasubramanian, C., Kumar, N., Agrawal, S., Ojha, R. P., and Maiti, S. (2008) Binding of berberine to human telomeric quadruplex - spectroscopic, calorimetric and molecular modeling studies. *FEBS J.* 275, 3971–3983.
- (10) Maiti, M., and Kumar, G. S. (2010) Polymorphic nucleic acid binding of bioactive isoquinoline alkaloids and their role in cancer. *J Nucleic Acids* 2010, DOI: 10.4061/2010/S93408.
- (11) Bhadra, Kakali, and Suresh, K. G. (2011) Interaction of berberine, palmatine, coralyn, and sanguinarine to quadruplex DNA: A comparative spectroscopic and calorimetric study. *Biochim. Biophys. Acta, Gen. Subj.* 1810, 485–496.
- (12) Liu, Y. Q., Zheng, B., Xu, X. J., and Yuan, G. (2010) Probing the binding affinity of small-molecule natural products to the G-quadruplex in C-myc oncogene by electrospray ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* 24, 3072–3075.
- (13) Li, W., Zhang, M., Zhang, J. L., Li, H. Q., Zhang, X. C., Sun, Q., and Qiu, C. M. (2006) Interactions of daidzin with intramolecular G-quadruplex. *FEBS Lett.* 580, 4905–4910.
- (14) Yang, S., Xiang, J., Yang, Q., Zhou, Q., Zhang, X., Li, Q., Tang, Y., and Xu, G. (2010) Distinct G-quadruplex structures of human telomeric DNA formed by the induction of sanguinarine and nitidine under salt-deficient condition. *Fitoterapia* 81, 1026–1032.
- (15) Bai, L. P., Hagihara, M., Jiang, Z. H., and Nakatani, K. (2008) Ligand binding to tandem G quadruplexes from human telomeric DNA. *ChemBioChem* 9, 2583–2587.
- (16) Petraccone, L., Garbett, N. C., Chaires, J. B., and Trent, J. O. (2010) An integrated molecular dynamics (MD) and experimental study of higher order human telomeric quadruplexes. *Biopolymers* 93, 533–548.
- (17) Hänsel, R., Löhr, F., Foldynova-Trantirkova, S., Bamberg, E., Trantirek, L., and 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.
- (18) Dai, J. X., Carver, M., and Yang, D. Z. (2008) Polymorphism of human telomeric quadruplex structures. *Biochimie* 90, 1172–1183.
- (19) Phan, A. T. (2010) Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS J.* 277, 1107–1117.
- (20) Wang, Y., and Patel, D. J. (1993) Solution structure of the human telomeric repeat D[Ag(3)(T(2)Ag(3))3] G-tetraplex. *Structure* 1, 263–282.
- (21) Dai, J. X., Punchihewa, C., Ambrus, A., Chen, D., Jones, R. A., and Yang, D. Z. (2007) Structure of the intramolecular human telomeric G-quadruplex in potassium solution: a novel adenine triple formation. *Nucleic Acids Res.* 35, 2440–2450.
- (22) Dai, J. X., Carver, M., Punchihewa, C., Jones, R. A., and Yang, D. Z. (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.
- (23) Maiti, M., Nandi, R., and Chaudhuri, K. (1983) The effect of Ph on the absorption and fluorescence-spectra of sanguinarine. *Photochem. Photobiol.* 38, 245–249.
- (24) Jones, R. R., Harkrader, R. J., and Southard, G. L. (1986) The effect of pH on sanguinarine iminium ion form. *J. Nat. Prod.* 49, 1109–1111.
- (25) Zhang, W. J., Ou, T. M., Lu, Y. J., Huang, Y. Y., Wu, W. B., Huang, Z. S., Zhou, J. L., Wong, K. Y., and Gu, L. Q. (2007) 9-Substituted berberine derivatives as G-quadruplex stabilizing ligands in telomeric DNA. *Bioorg. Med. Chem.* 15, 5493–5501.
- (26) Maiti, M., and Chaudhuri, K. (1981) Interaction of berberine chloride with naturally occurring deoxyribonucleic acids. *Indian J. Biochem. Biophys.* 18, 245–250.
- (27) Tripathi, A. N., Chauhan, L., Thankachan, P. P., and Barthwal, R. (2007) Quantum chemical and nuclear magnetic resonance spectral studies on molecular properties and electronic structure of berberine and berberrubine. *Magn. Reson. Chem.* 45, 647–655.
- (28) Marek, R., Tousek, J., Dostal, J., Slavik, J., Dommissie, R., and Sklenar, V. (1999) H-1 and C-13 NMR study of quaternary benzo[c]phenanthridine alkaloids. *Magn. Reson. Chem.* 37, 781–787.
- (29) Dixon, I. M., Lopez, F., Esteve, J. P., Tejera, A. M., Blasco, M. A., Pratviel, G., and Meunier, B. (2005) Porphyrin derivatives for telomere binding and telomerase inhibition. *ChemBioChem* 6, 123–132.
- (30) Buurma, N. J., and Haq, I. (2008) Calorimetric and spectroscopic studies of Hoechst 33258: Self-association and binding to non-cognate DNA. *J. Mol. Biol.* 381, 607–621.
- (31) Parkinson, G. N., Ghosh, R., and Neidle, S. (2007) Structural basis for binding of porphyrin to human telomeres. *Biochemistry* 46, 2390–2397.
- (32) Parkinson, G. N., Cuenca, F., and Neidle, S. (2008) Topology conservation and loop flexibility in quadruplex-drug recognition: Crystal structures of inter- and intramolecular telomeric DNA quadruplex-drug complexes. *J. Mol. Biol.* 381, 1145–1156.
- (33) Campbell, N. H., Parkinson, G. N., Reszka, A. P., and Neidle, S. (2008) Structural basis of DNA quadruplex recognition by an acridine drug. *J. Am. Chem. Soc.* 130, 6722.
- (34) McGhee, J. D., and Von Hippel, P. H. (1974) Theoretical aspects of DNA-protein interactions: Co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice. *J. Mol. Biol.* 86, 469–489.
- (35) Hwang, T. L., and Shaka, A. J. (1995) Water suppression that works - Excitation sculpting using arbitrary wave-forms and pulsed-field gradients. *J. Magn. Reson. Ser. A* 112, 275–279.
- (36) Bax, A., and Davis, D. G. (1985) Practical aspects of two-dimensional transverse NOE spectroscopy. *J. Magn. Reson.* 63, 207–213.
- (37) Schleucher, J., Schwendinger, M., Sattler, M., Schmidt, P., Schedletsky, O., Glaser, S. J., Sorensen, O. W., and Griesinger, C. (1994) A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed-field gradients. *J. Biomol. NMR* 4, 301–306.
- (38) Kay, L. E., Keifer, P., and Saarinen, T. (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* 114, 10663–10665.
- (39) Palmer, A. G., Cavanagh, J., Wright, P. E., and Rance, M. (1991) Sensitivity improvement in proton-detected 2-dimensional heteronuclear correlation NMR-spectroscopy. *J. Magn. Reson.* 93, 151–170.
- (40) Sklenar, V., and Bax, A. (1987) Spin-Echo water suppression for the generation of pure-phase two-dimensional NMR-spectra. *J. Magn. Reson.* 74, 469.
- (41) Piotto, M., Saudek, V., and Sklenar, V. (1992) Gradient-Tailored Excitation for Single-Quantum NMR-Spectroscopy of Aqueous-Solutions. *J. Biomol. NMR* 2, 661–666.

- (42) Sklenar, V., Piotto, M., Leppik, R., and Saudek, V. (1993) Gradient-Tailored Water Suppression for ^1H - ^{15}N HSQC Experiments Optimized to Retain Full Sensitivity. *J. Magn. Reson. Ser. A* 102, 241–245.
- (43) Goddard, T. D., Kneller, D. G. (2007) SPARKY 3, University of California, San Francisco.
- (44) *Maestro v 8.5* (2007) Schrödinger, LLC, New York, NY.
- (45) *Glide v. 4.5* (2007) Schrödinger, LLC, New York, NY.
- (46) *MacroModel v. 9.6* (2008) Schrödinger, L.L.C., New York, NY.
- (47) Bayly, C. I., Cieplak, P., Cornell, W. D., and Kollman, P. A. (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges - the Resp model. *J. Phys. Chem.* 97, 10269–10280.
- (48) Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery Jr., J. A., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, J. M., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, Ö., Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J. (2009) *Gaussian 09*, Revision A.01 ed., Gaussian, Inc., Wallingford, CT.
- (49) Case, D. A., Darden, T. A., Cheatham, T. E., III, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Merz, K. M., Pearlman, D. A., Crowley, M., Walker, R. C., Zhang, W., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Wong, K. F., Paesani, F., Wu, X., Brozell, S., Tsui, V., Gohlke, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, G., Beroza, P., Mathews, D. H., Schafmeister, C., Ross, W. S., and Kollman, P. A. (2006) *AMBER 9*, University of California, San Francisco, CA.
- (50) Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G. M., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J. M., and Kollman, P. (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* 24, 1999–2012.
- (51) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935.
- (52) Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.* 103, 8577–8593.
- (53) Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical-integration of cartesian equations of motion of a system with constraints - molecular-dynamics of N-alkanes. *J. Comput. Phys.* 23, 327–341.
- (54) Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., Dinola, A., and Haak, J. R. (1984) Molecular-dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690.
- (55) Dominguez, C., Boelens, R., and Bonvin, A. M. J. J. (2003) HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* 125, 1731–1737.
- (56) van Dijk, M., van Dijk, A. D. J., Hsu, V., Boelens, R., and Bonvin, A. M. J. J. (2006) Information-driven protein-DNA docking using HADDOCK: it is a matter of flexibility. *Nucleic Acids Res.* 34, 3317–3325.
- (57) Schüttelkopf, A. W., and van Aalten, D. M. F. (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 60, 1355–1363.
- (58) Linge, J. P., Williams, M. A., Spronk, C. A. E. M., Bonvin, A. M. J. J., and Nilges, M. (2003) Refinement of protein structures in explicit solvent. *Proteins: Struct., Funct., Bioinf.* 50, 496–506.
- (59) Jorgensen, W. L., and Tirado-Rives, J. (1988) The OPLS force field for proteins. energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* 110, 1657–1666.
- (60) Pranata, J., Wierschke, S. G., and Jorgensen, W. L. (1991) OPLS potential functions for nucleotide bases - relative association constants of hydrogen-bonded base-pairs in chloroform. *J. Am. Chem. Soc.* 113, 2810–2819.
- (61) Nozinovic, S., Fürting, B., Jonker, H. R. A., Richter, C., and Schwalbe, H. (2010) High-resolution NMR structure of an RNA model system: the 14-mer cUUCGg tetraloop hairpin RNA. *Nucleic Acids Res.* 38, 683–694.

6.2 Research article: A nucleus-imaging probe that selectively stabilizes a minor conformation of *c-MYC* G-quadruplex and down-regulates *c-MYC* transcription in human cancer cells

Panda D., Debnath M., Mandal S., Bessi I., Schwalbe H., Dash J.
Scientific Reports, 2015, 19; 5: 13183.

The effects of a fluorescent bis-triazolylcarbazole derivative (BTC-f) on the *c-MYC* expression and its potential application as nucleus imaging probe for G-quadruplexes are discussed in this article.

A series of bis-triazolylcarbazole derivatives (BTC) was prepared using a modular synthesis approach. The derivative BTC-f was found to bind to different promoter quadruplexes, with highest affinity for *c-MYC* and selectivity over duplex DNA. NMR titration of the *c-MYC* quadruplex with BTC-f revealed that the ligand is able to bind and stabilize a minor populated conformer *via* conformational selection. The fluorescence enhancement detected upon binding to the promoter G-quadruplexes allows BTC-f to be used as a selective "turn-on" probe for detection of quadruplex over duplex DNA in living cells. The ligand BTC-f was able to down-regulate the *c-MYC* expression and to induce apoptosis in cancer cells, possibly by stabilization of the *c-MYC* G-quadruplex.

The design and the synthesis of the bis-triazolylcarbazole ligands, the FRET studies as well as all the biological essays were carried out in the group of Prof. Dash (IISER Kolkata, India). The author of this thesis performed the NMR titration experiments, analysed the data, contributed to elucidate the interaction mode of *c-MYC* quadruplex with BTC and wrote the corresponding paragraph in the article.

SCIENTIFIC REPORTS

OPEN

A Nucleus-Imaging Probe That Selectively Stabilizes a Minor Conformation of *c-MYC* G-quadruplex and Down-regulates *c-MYC* Transcription in Human Cancer Cells

Received: 07 April 2015
Accepted: 17 July 2015
Published: 19 August 2015

Deepanjan Panda¹, Manish Debnath¹, Samir Mandal¹, Irene Bessi², Harald Schwalbe² & Jyotirmayee Dash¹

The *c-MYC* proto-oncogene is a regulator of fundamental cellular processes such as cell cycle progression and apoptosis. The development of novel *c-MYC* inhibitors that can act by targeting the *c-MYC* DNA G-quadruplex at the level of transcription would provide potential insight into structure-based design of small molecules and lead to a promising arena for cancer therapy. Herein we report our finding that two simple bis-triazolylcarbazole derivatives can inhibit *c-MYC* transcription, possibly by stabilizing the *c-MYC* G-quadruplex. These compounds are prepared using a facile and modular approach based on Cu(I) catalysed azide and alkyne cycloaddition. A carbazole ligand with carboxamide side chains is found to be microenvironment-sensitive and highly selective for “turn-on” detection of *c-MYC* quadruplex over duplex DNA. This fluorescent probe is applicable to visualize the cellular nucleus in living cells. Interestingly, the ligand binds to *c-MYC* in an asymmetric fashion and selects the minor-populated conformer via conformational selection.

The *c-MYC* oncogene is overexpressed in a broad spectrum of human malignancies and emerged as a potential therapeutic target for cancer treatment¹. The overexpression of *c-MYC* in hepatic cells is frequently associated with the development of hepatocellular carcinoma². Small molecule mediated inhibition of *c-MYC* imparts growth arrest in liver cancer cells and even down-regulates the human telomerase reverse transcriptase (hTERT) activity^{3,4}. Therefore *c-MYC* is an attractive target in developing new therapies for hepatocellular carcinoma. Transcription of *c-MYC* is primarily regulated by a 27 base guanine-rich sequence present within the nuclease hypersensitivity element III1 (NHE III1)⁵. This sequence, located -142 to -115bp upstream of the P1 promoter of *c-MYC* oncogene, exists in equilibrium between transcriptionally active forms (double helical and single stranded) and a silenced form, which is able to fold into a G-quadruplex structure⁶. The 27-mer sequence responsible of *c-MYC* regulation contains five guanine runs and it has been shown that in K⁺-containing solution only the four consecutive 3' G-runs are involved in the formation of the major G-quadruplex structure which causes the gene transcriptional silencing⁷⁻⁸. However, the major G-quadruplex formed under superhelicity conditions involves the four

¹Department of Organic Chemistry, Indian Association for the Cultivation of Science, Jadavpur, Kolkata-700032, India. ²Institute of Organic Chemistry and Chemical Biology, Goethe University Frankfurt and Centre for Biomolecular, Magnetic Resonance, Max-von-Laue Strasse 7, 60438, Frankfurt am Main, Germany. Correspondence and requests for materials should be addressed to J.D. (email: ocjd@iacs.res.in)

consecutive 5' G-runs^{9,10}. Various G-quadruplex structures derived from different G-rich tracts of the *c-MYC* regulatory element NHE III1 have been reported^{11–13}. The pioneering work of Siddiqui-Jain *et al.* showed that small molecules stabilizing the *c-MYC* G-quadruplex can reduce *c-MYC* transcription in cancer cells¹⁴. Since then, several classes of small molecules, which can bind and stabilize *c-MYC* quadruplex have been developed^{15–27} and few of them have been structurally characterized in complex with *c-MYC* G-quadruplex by NMR^{12,24}. However, only a few molecules, for example, cationic porphyrins²⁰, quindolines²¹ and metal complexes²⁵ have been taken forward to cellular system to exert their desired biological effect¹⁵. In most cases, either the ligand has been prepared using multistep syntheses with low overall yields making it not readily accessible or the mode of binding to the *c-MYC* quadruplex remain unknown. Therefore, it is important to devise easily synthesizable and cost-effective *c-MYC* stabilizers which can exhibit tailored antiproliferative activities in cancer cells.

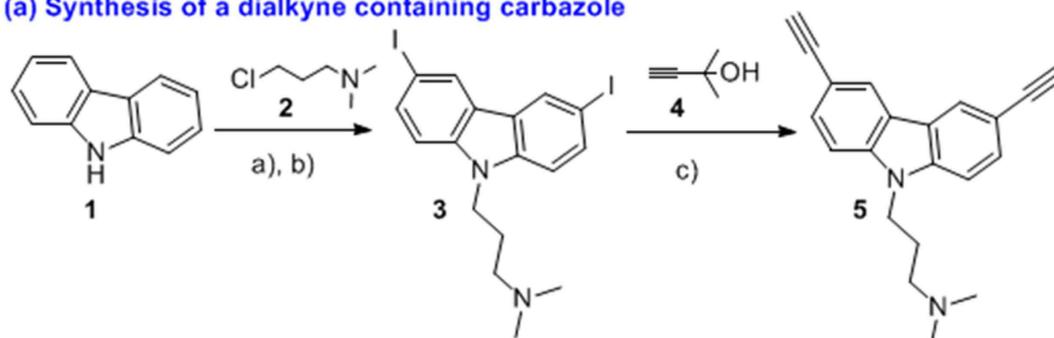
Carbazole derivatives exhibit a wide range of pharmacological activities. A few carbazole derivatives have been reported to bind to *h-TELO* G-quadruplex^{27–28}. Herein, we delineate a modular synthetic access to novel bis-triazolyl carbazole derivatives²⁹ as potent “turn on” G-quadruplex probes, which bind to *c-MYC* quadruplex *via* conformational selection with the potential to down-regulate *c-MYC* transcription in hepatocellular carcinoma cells.

Results

Modular synthesis of bis-triazolyl carbazole ligands. For the synthesis, we prepared a carbazole dialkyne **5** from the commercially available carbazole **1**. The iodination³⁰ of **1** followed by N-alkylation with 3-dimethylaminopropylchloride (**2**) afforded the diiodo compound **3**. Sonogashira coupling of **3** with 3-methyl butynol **4** followed by removal of the acetone group afforded the dialkyne **5** in 90% overall yield for the two steps. Carbazole dialkyne **5** was treated with the azides **6a–g** (see Supplementary Information) using catalytic CuSO₄, sodium ascorbate in *t*-BuOH/H₂O (3:1)³¹ under microwave conditions at 70 °C for 4 h to give the corresponding bis-triazolylcarbazole derivatives **BTC a–g** in high yields (Fig. 1 and Supplementary Chart S1).

Bis-triazolylcarbazoles selectively stabilize G-quadruplex over duplex DNA. Förster Resonance Energy Transfer (FRET) melting analysis^{32,33} was employed to determine the stabilization induced by carbazole derivatives (**5** and **BTC a–g**) by measuring the change in melting temperature (ΔT_m) of the dual labeled (5'-FAM and 3'-TAMRA) G-rich oncogenic promoter sequences (*c-MYC*, *c-KIT1* and *c-KIT2*) and a control *ds DNA*. Ligand **BTC f** containing the three terminal-NMe₂ groups exhibited maximum stabilization for quadruplexes at a ligand concentration of 1 μ M. However, the starting dialkyne-carbazole **5** hardly altered the T_m values of the quadruplexes. Ligands **BTC a–c** lacked the two cationic side chains and displayed moderate G-quadruplex stabilization. When the two cationic side chains of the carbazole were substituted by two anionic –COOH groups (ligand **BTC e**), a diminished stabilization for the quadruplex sequences was observed. These observations collectively suggest that changing the end substituents of the carbazole ligands has a profound impact on G-quadruplex stabilization (Fig. 2a and Supplementary Table S1). FRET competitive experiments were performed to evaluate the selectivity of **BTC f** for G-quadruplex over duplex DNA (Fig. 2b). FRET melting of 200 nM dual labeled *c-MYC*, *c-KIT1* and *c-KIT2* quadruplexes were performed with 1 μ M **BTC f** in the presence of different concentrations of competitor *ds DNA* (0, 0.2, 1.0, 2.0, 20 μ M). **BTC f** appeared to be highly selective towards the quadruplexes by maintaining high ΔT_m values for the quadruplexes even in the presence of 10 mol equivalent excess of *ds DNA*. At 100 mol equivalent excess of *ds DNA*, only a small change in ΔT_m values for the quadruplexes was observed. These results indicate that *ds DNA* does not significantly interfere with the binding of **BTC f** to the quadruplexes. Next, concentration-dependent FRET melting experiments of the promoter quadruplexes and *ds DNA* were performed. The melting profiles at various concentrations of **BTC f** demonstrate a dose-dependent increase in the values of ΔT_m for the quadruplexes (Fig. 2c and Supplementary Table S2). Ligand **BTC f** showed the highest ΔT_m for promoter gene quadruplexes that can be measured by this method. It showed a ΔT_m value of 22.7 ± 1.4 K (i.e. a T_m of 93 °C) for *c-MYC* at 100 nM, 39.4 ± 2.5 K (i.e. a T_m of 93 °C) for *c-KIT1* at 750 nM and 23.2 ± 1.6 K (i.e. a T_m of 93 °C) for *c-KIT2* at 500 nM ligand concentrations. These results revealed that 5–7 fold higher concentrations of **BTC f** are required for the *c-KIT1* and *c-KIT2* to achieve maximum stabilization potential (ΔT_m) compared to the *c-MYC* quadruplex. To investigate whether **BTC f** could induce the formation of the G-quadruplex, FRET melting assay was performed in the absence of K⁺ ion. Interestingly, **BTC f** could stabilize G-quadruplex formation in the absence of any added K⁺ ion, pH 7.4 (Supplementary Fig. S1). Similar to the results of FRET melting in the presence of K⁺, lower concentrations (250 nM) of **BTC f** were required to obtain the highest ΔT_m for *c-MYC* in the absence of K⁺. Higher concentrations of **BTC f** are required for achieving maximum ΔT_m for the *c-KIT1* (1 μ M) and *c-KIT2* (750 nM) quadruplexes. **BTC f** did not significantly change the melting temperature of *ds DNA* even at higher concentrations.

BTC f as an environmentally sensitive “turn-on” probe for *c-MYC* G-quadruplex. The binding characteristics of **BTC f** with the G-quadruplex forming sequences (*c-MYC*, *c-KIT1* and *c-KIT2*) and duplex DNA (*ds DNA*) were investigated using fluorescence spectroscopy (Fig. 3 and Supplementary Fig. S2). The emission spectra of **BTC f** displayed a maxima at 522 nm (quantum yield, $\Phi = 0.054$, abs,

(a) Synthesis of a dialkyne containing carbazole

a) KI, KIO₃, acetic acid, water, 80 °C, 48 h, 65%, b) **2**, THF, NaH, 75 °C, 26 h, 88%
 c) (i) **4**, PdCl₂(PPh₃)₂, CuI, Et₃N, rt, 12 h, (ii) KOH, toluene, 110 °C, 12 h, 90%

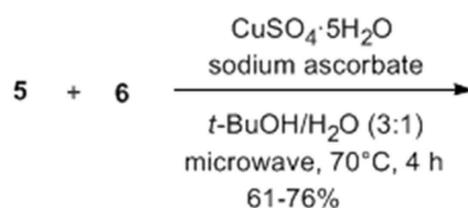
(b) Azides used in the "Click reaction"

Figure 1. Modular synthesis of bis-triazolylcarbazole ligands (BTC).

λ_{\max} = 260, 290 nm), when excited at 290 nm in 100 mM Tris•HCl buffer, at pH 7.4 containing 100 mM KCl. Upon titration with the promoter quadruplexes (0–6 equiv.), **BTC f** (1.0 μ M) showed an increase in fluorescence (2–3.3 fold) along with a blue shift (14–20 nm). The degree of fluorescence enhancement was the highest for *c-MYC* (3.3 fold) compared to the *c-KIT1* (2.0 fold) and *c-KIT2* (2.0 fold) quadruplexes. The blue shift was the largest for the *c-MYC* (20 nm) compared to the *c-KIT1* (15 nm) and *c-KIT2* (14 nm). Almost negligible fluorescence enhancement was observed (Fig. 3c) when **BTC f** was titrated with the *ds DNA* (0–6 equiv.). The dissociation constants (K_D) calculated from the fluorescence spectra were found to be in the low micromolar range (0.3–1.38 μ M). Ligand **BTC f** shows a 5-fold selectivity for *c-MYC* with a K_D value of ($K_D = 300 \pm 15$ nM), over *c-KIT1* ($K_D = 1.38 \pm 0.07$ μ M) and *c-KIT2* ($K_D = 1.37 \pm 0.06$ μ M) quadruplexes. These results are consistent with the FRET melting data that **BTC f** shows selectivity for the *c-MYC* compared to the *c-KIT1* and *c-KIT2* quadruplexes. It is worth mentioning that **BTC f** exhibited a strong blue shift of 25 nm with a 5.5 fold increase in fluorescence upon titration with the *c-MYC* in the absence of K⁺ (Fig. 3b).

Next, the binding behaviour of **BTC c** with the *c-MYC* quadruplex was investigated (Supplementary Fig. S3a-b). The emission spectra of **BTC c** showed an emission maximum at 402 nm with a comparatively lower quantum yield (quantum yield, $\Phi = 0.044$, abs, $\lambda_{\max} = 256, 342$ nm) along with two weak shoulders, when excited at 330 nm in 100 mM KCl and 100 mM Tris•HCl buffer, at pH 7.4. When the fluorimetric

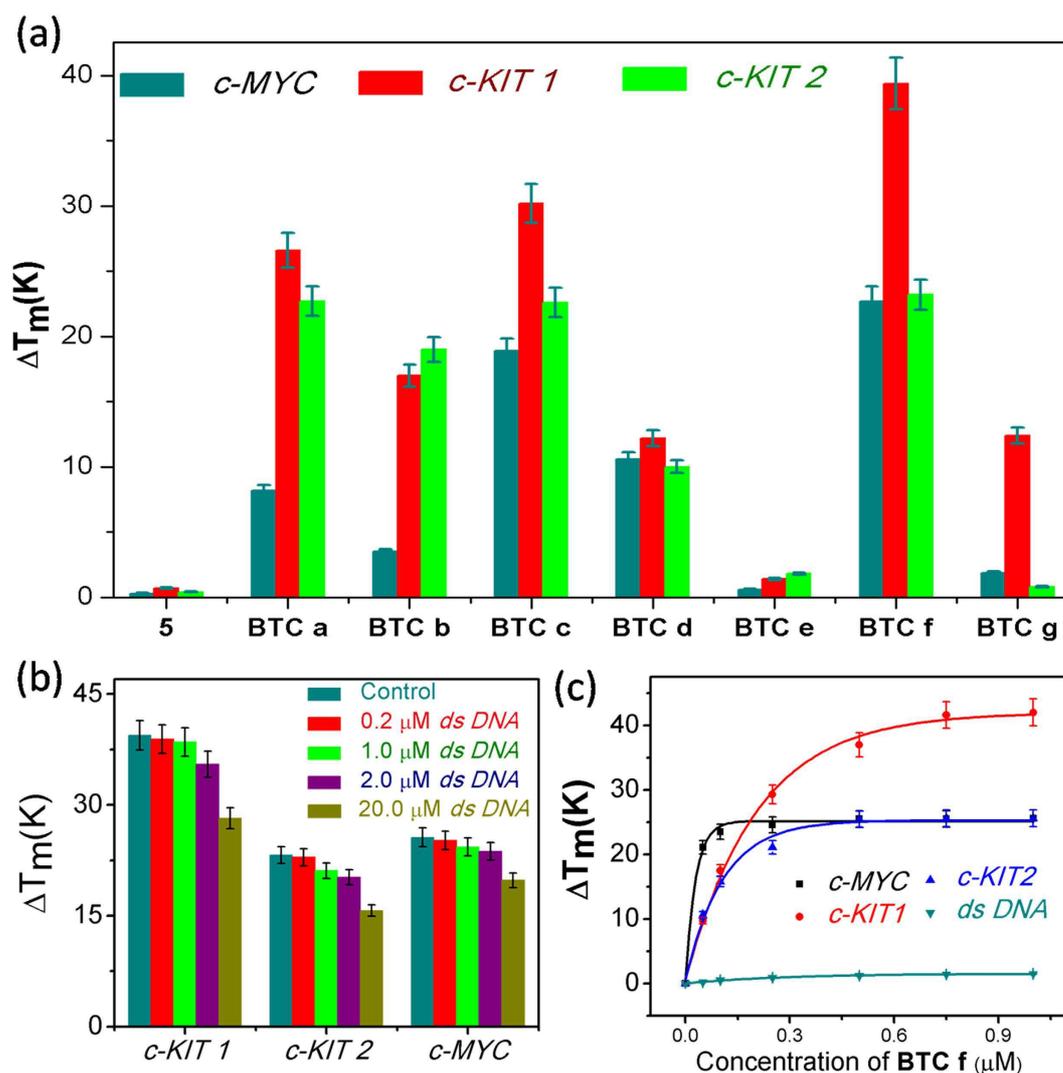


Figure 2. (a) FRET stabilization potential of carbazole derivatives **5** and **BTC a-g** (1.0 μM) upon binding to *c-MYC*, *c-KIT1* and *c-KIT2* quadruplex, T_m ($^{\circ}\text{C}$) for *c-MYC* = (70.9 ± 1.2) , *c-KIT1* = (53.0 ± 1.7) , *c-KIT2* = (70.0 ± 1.4) , *ds DNA* = (61.2 ± 2.1) . (b) FRET competition assay of **BTC f** (1.0 μM) for G-quadruplexes (200 nM) in the presence of duplex DNA (200 nM, 1.0 μM , 2.0 μM and 20.0 μM). (c) Thermal shift profiles for **BTC f** upon stabilizing to quadruplexes and duplex DNA; buffer: 50 mM potassium cacodylate, pH 7.4.

titration of **BTC c** (1 μM) was carried out with increasing concentration of the *c-MYC* (0–6equiv.), only a slight increase in fluorescence intensity was observed. The difference in the fluorescence spectra between **BTC c** and **BTC f** may be attributed to the presence of an extended π -electron delocalization between the lone pair of electrons on triazole nitrogen atom of the donor and the acceptor carbonyl group in **BTC f**. These results indicate that **BTC f** is an effective light-up probe for the *c-MYC* quadruplex. In good agreement with the FRET melting results, **BTC f** that exhibited a higher stabilization potential showed several fold higher binding affinity for the *c-MYC* ($K_D = 300 \pm 15$ nM) compared to **BTC c** ($K_D > 20$ μM).

In order to understand the observed blue shift upon interaction of **BTC f** with quadruplexes, the effect of environment polarity was analysed by recording the emission spectra of **BTC f** in various solvent environments such as ethyl acetate, DMF, DMSO, ethanol, methanol and water (Supplementary Fig. S4a). The observed emissions (λ_{max}^{em}) were plotted against Reichardt's $E_T(30)$ polarity parameter³⁴ (Supplementary Fig. S4b) to give a linear correlation ($y = 4.03x + 267.3$, $r^2 = 0.9292$) over the range of the solvents. The Stokes shift (S) was calculated from the maximum wavelength (λ_{max}) value of the excitation and emission spectra (Supplementary Table S4). Both λ_{max}^{em} and S values showed their strong dependence on the polarity of the medium. We observed that the emission maximum of **BTC f** is blue-shifted from 522 to 410 nm and the fluorescence intensity is simultaneously enhanced as the solvent polarity decreases from water to ethylacetate. These observations suggest that the enhancement in the fluorescence intensity along with

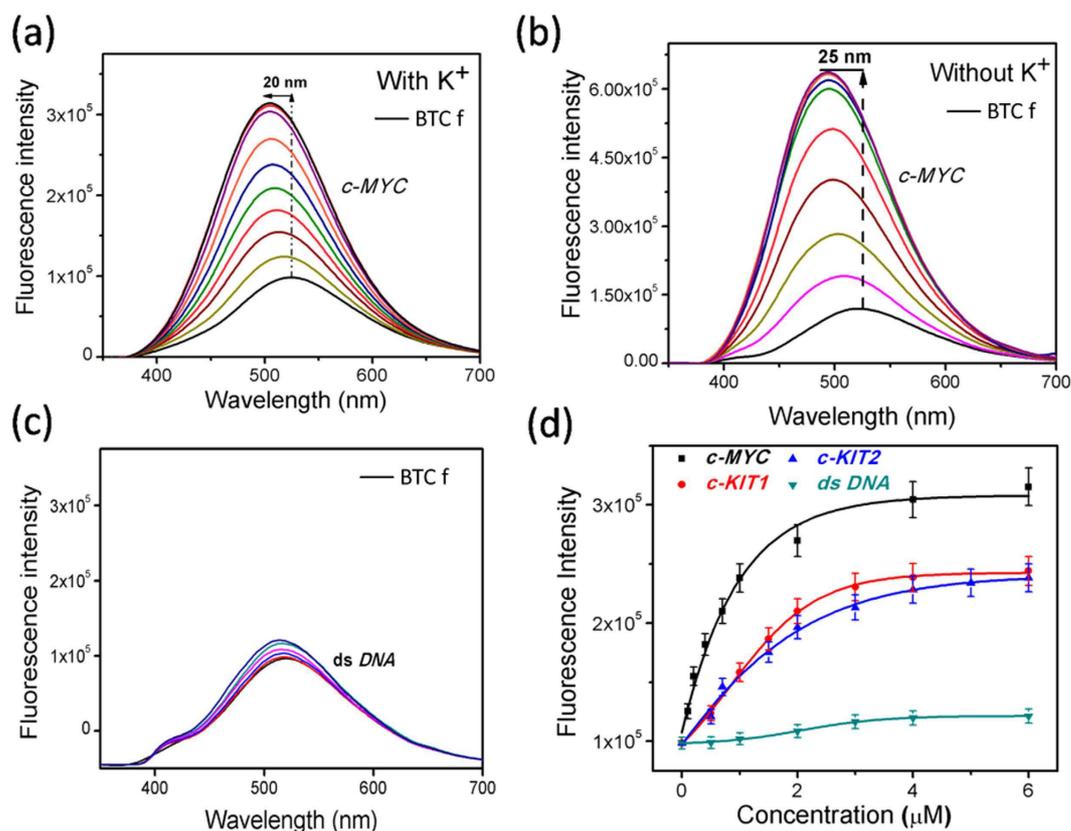


Figure 3. Fluorimetric titration spectra of BTC f (1.0 μM) with (0–6 eq) of (a) *c-MYC* in 100 mM Tris•HCl buffer at pH 7.4 containing 100 mM KCl, (b) *c-MYC* in 100 mM Tris•HCl buffer at pH 7.4 without additional KCl and (c) *ds DNA* in 100 mM Tris•HCl buffer at pH 7.4 containing 100 mM KCl. (d) Fluorescence intensity profiles of BTC f (1.0 μM) upon step-wise addition of quadruplex sequences (0–6 eq); titrations were performed in 100 mM Tris•HCl buffer at pH 7.4 containing 100 mM KCl.

the blue-shift in the emission spectra of the BTC f with the quadruplexes is a consequence of the proximity of BTC f to the non-polar hydrophobic regions³⁵ of the G-quadruplexes thereby manifesting the high affinity of BTC f towards the G-quadruplex structures.

BTC acquires an induced CD signal upon interaction with *c-MYC* G-quadruplex. The circular dichroism (CD) spectrum of *c-MYC* quadruplex sequence showed a positive peak at 260 nm and a negative peak around 240 nm, in the absence of K⁺, which is characteristic of a parallel conformation^{36–38} (Supplementary Fig. S5). As potassium stabilizes G-quadruplex structure, the molar ellipticity of both positive and negative peaks of *c-MYC* was enhanced in K⁺ containing buffer (Supplementary Fig. S5). The incremental addition of BTC f to the *c-MYC* resulted in a slight decrease in the positive band at 260 nm in the presence and absence of K⁺. This suggests that BTC f binds to the parallel quadruplex structure of *c-MYC* and does not disrupt the structure. In addition, BTC f exhibited a positive induced circular dichroism (ICD) in the absorbing 290–350 nm range (centered at 322 nm) in the presence and absence of K⁺. It is worth mentioning that, no such ICD signal was detected upon binding of BTC f to *c-KIT 1* and *c-KIT 2* quadruplexes (Supplementary Fig. S5c-d), which indicates that BTC f can discriminate between the promoter quadruplexes.

UV/Vis spectroscopy indicated that BTC f exhibits hypochromism upon addition of the *c-MYC* (0–0.4 equiv.) and the peaks at 260–290 nm overlap with the *c-MYC* at higher concentration (Supplementary Fig. S6). The observed hypochromism in UV spectra suggests strong stacking interaction between aromatic chromophores of BTC f with the G-quadruplex base pairs and indicates possible binding of these ligands into the nonpolar site of the *c-MYC* quadruplex. The red shifted ICD band at 322 nm in CD spectra may arise due to the binding of the optically inactive BTC f inside the asymmetric microenvironment of the *c-MYC* quadruplex³⁹. CD spectroscopic analysis further reveals that BTC f bound *c-MYC* structure is similar in the presence and absence of K⁺.

Bis-triazolylcarbazoles selectively bind to a minor conformation of *c-MYC* quadruplex. NMR titrations show that the ligand BTC f is strongly interacting with the *c-MYC* sequence, as suggested

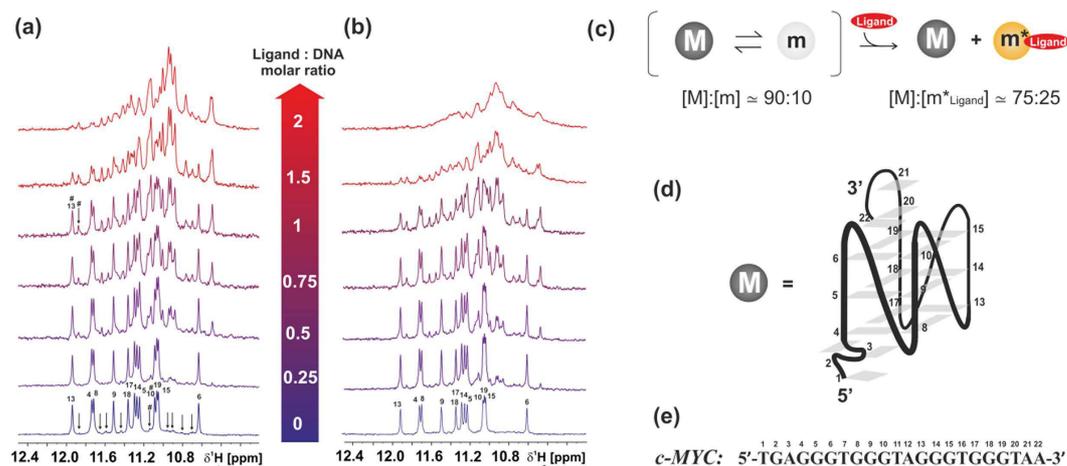


Figure 4. Imino region of 100 μM *c*-MYC in the presence of BTC f at different [ligand]:[DNA] molar ratio, (a) in 25 mM Tris-HCl buffer at pH 7.4 containing 100 mM KCl, 90% $\text{H}_2\text{O}/10\%$ D_2O (b) in 25 mM Tris-HCl buffer at pH 7.4 without additional KCl, 90% $\text{H}_2\text{O}/10\%$ D_2O . Titrations were performed at 298 K, 600 MHz. Signals from the major conformation of *c*-MYC are labeled according to the numbering shown in panel (e), while signals from the binding-competent minor conformation are marked with arrows.

(c) Scheme of the proposed binding mechanism with estimated population of each state. Major conformation, minor conformation in the free-form and minor conformation in the ligand-bound form are indicated with M, m and m^*_{Ligand} , respectively. Population of the states in the absence of the ligand was estimated from the ratio of the integrals of signals marked in the spectrum with #, while population of the states in presence of ligand was estimated at a [ligand]:[DNA] molar ratio of 1, from the ratio of the integrals of signals marked in the spectrum with #. (d) Folding of the major conformation (M) of *c*-MYC determined by Ambrus *et al.* [PDB code: 1XAV]¹³. (e) *c*-MYC sequence used for NMR binding titration.

by the significant changes of the imino pattern profile upon addition of the ligand (Fig. 4). Figure 4a shows that upon addition of BTC f, a set of minor peaks (already observable in the DNA free form and indicated with arrows) becomes more intense, while the chemical shift and the line width of the imino protons of the major conformation are not perturbed. At a [ligand]:[DNA] molar ratio of 2, the intensity of the peaks of the major conformation is reduced, compared to the intensity of the peaks of the ligand-bound conformation. We propose that BTC f binds to the *c*-MYC via conformational selection: out of an ensemble of DNA conformers in equilibrium with each other (major conformation-Fig. 4d and one or several minor conformations), the ligand selects the minor-populated conformer and binds to it in a specific manner. Interestingly, BTC f binds to the minor conformer, most likely a parallel stranded structure with different capping structures¹³, and upon ligand binding, the equilibrium of the free form $\text{DNA}^{\text{major}}/\text{DNA}^{\text{minor}}$ conformational equilibrium is accordingly shifted. Moreover, upon ligand addition we also observe the appearance of new signals in the imino region (e.g., at 10.5 ppm) suggesting that the ligand BTC f can induce conformational changes in the structure of the binding-competent minor conformation. The proposed binding mechanism is summarized in Fig. 4c. The major conformation (M) is not binding-competent and the binding-competent minor conformation (m) undergoes small conformational changes upon ligand binding (m^*_{Ligand}). NMR titration of BTC f in the absence of KCl (Fig. 4b) shows that the ligand bound conformation of *c*-MYC resembles the one observed in the presence of 100 mM KCl. These results support FRET melting, fluorescence and CD studies that BTC f can stabilize the *c*-MYC quadruplex in the absence and presence of potassium. NMR titration of *c*-MYC with ligand BTC c, which lacks the two cationic side chains showed line broadening with no significant chemical shift perturbation of the imino proton signals of the DNA major conformation (Supplementary Fig. S7a). Similar to the [BTC f]:[*c*-MYC] complex, upon addition of BTC c, a new set of peaks was observed, which corresponds to one or more minor conformers of the *c*-MYC quadruplex. The NMR data suggest that these ligands interact with and stabilize a minor conformation of *c*-MYC quadruplex.

Bis-triazolyl carbazoles inhibit the growth of cancer cells. To characterize the activity in biological systems, the antiproliferative activities of this class of compounds BTC a–g were determined in human hepatocellular liver carcinoma HepG2 cells using MTT assays⁴⁰ (Supplementary Fig. S8 and Table S5). The results indicated that BTC f potently inhibits the growth of cancer cells at low micromolar concentrations with an IC_{50} value of $\sim 4.3 \pm 0.69 \mu\text{M}$. BTC c ($\text{IC}_{50} = 11.38 \pm 1.54 \mu\text{M}$) displays a somewhat lower antiproliferative activity than BTC f. However, BTC d–e and BTC g do not show strong activity against HepG2 cells (Supplementary Table S5). We then evaluated the growth inhibitory activity of BTC f in two other human cancer cell lines such as breast carcinoma (MCF-7) and colon carcinoma (HCT

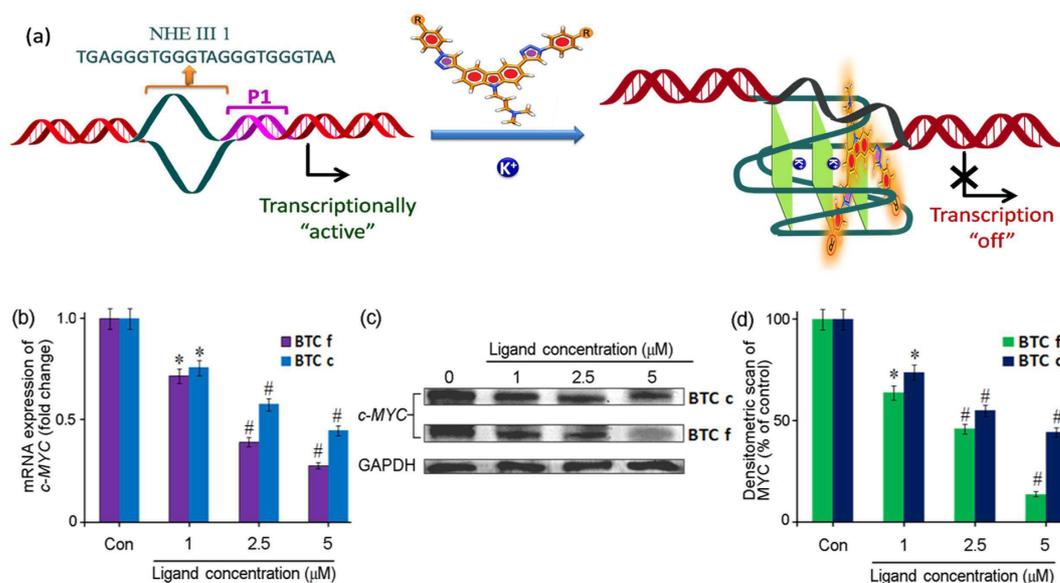


Figure 5. Effect of BTC c and f on the expression of *c-MYC* protein in human cancer cells: (a) Schematic representation of ligand induced transcriptional down-regulation of *c-MYC* gene. (b) Determination of transcriptional regulation of *c-MYC* mRNA in the presence of increasing concentration of BTC c or BTC f in cancer cells by qRT-PCR. qRT-PCR was performed and quantified by comparative threshold method using GAPDH as housekeeping gene. (c) Immunoreactive bands of MYC protein were analyzed by Western blot. (d) Densitometric analyses of immunoblots showing concentration dependent reduced level of MYC protein. Results are representative of five independent experiments with two replicates. The data are shown as mean \pm SD. * $P < 0.05$, # $P < 0.01$, versus untreated cancer cells.

116), and a normal cell line (C2C12, mouse myoblast). BTC f exerts growth inhibition against human cancer cells, with IC_{50} values of $6.8 \pm 0.84 \mu\text{M}$ and $8.2 \pm 1.2 \mu\text{M}$ for MCF 7 and HCT 116 cells respectively (Supplementary Fig. S8), but importantly it is less toxic to the model normal cell line ($IC_{50} > 45 \mu\text{M}$ for C2C12, data not shown). It is worth noting that the main core carbazolidialkyne **5** does not show cytotoxicity towards normal as well as cancer cells (data not shown). These results indicate that the triazole ring and the side chains contribute towards the antiproliferative activities of the carbazole derivatives in cancer cells. Among the cancers cells (HepG2, MCF-7 and HCT 116), BTC f is the most potent in HepG2 cells. It has been reported that the expression *c-MYC* gene is elevated in liver carcinoma HepG2 cells compared to the normal human liver cells². Biophysical analysis suggested that BTC f shows high specificity for the *c-MYC* G-quadruplex over duplex DNA. This encouraged us to evaluate the molecular mechanism underlying the anti-proliferative effect of BTC f in HepG2 cells. For a comparison, the effect of the less potent *c-MYC* quadruplex binding ligand BTC c with a reduced antiproliferative activity towards HepG2 cells was also investigated.

Bis-triazolylcarbazoles down-regulate *c-MYC* expression. Ligands BTC c and BTC f were evaluated for their effect on transcriptional regulation of *c-MYC* in HepG2 cells (Fig. 5a). Total mRNA was isolated from HepG2 cells after treatment with varying concentrations (1.0, 2.5 and 5.0 μM) of BTC c and BTC f for 24 h. The level of *c-MYC* mRNA was quantified using quantitative real-time polymerase chain reaction (qRT-PCR) and the gene expression was normalized relative to the expression of a constitutively expressed house-keeping gene, glyceraldehyde-3-phosphate dehydrogenase (GAPDH). Analysis of the qRT-PCR data revealed that both BTC c and BTC f are able to reduce the level of *c-MYC* mRNA in a dose dependent manner (Fig. 5b). At 5.0 μM concentration, BTC f reduces the level of *c-MYC* mRNA by $75 \pm 3\%$ relative to the control. Compared to that, BTC c at 5.0 μM leads to a suppression of *c-MYC* mRNA level by $48 \pm 3\%$ of the control suggesting its relatively lower efficacy than BTC f. The GAPDH mRNA is equally expressed in control, BTC c and BTC f treated HepG2 cells, which confirms that the reduction of mRNA level due to BTC c and BTC f is *c-MYC* gene-specific. To investigate that the observed reduction in *c-MYC* mRNA levels leads to an inhibition of MYC protein, we have employed western blotting using anti-MYC antibody (Fig. 5c,d). Protein levels were measured for the *c-MYC* and the housekeeping gene GAPDH in HepG2 cells after treatment with BTC c and BTC f for 24 h at final concentrations of 1.0, 2.5 and 5.0 μM . Both BTC c and BTC f exhibit reduced expression of MYC protein in a dose dependent manner compared to the untreated HepG2 cells (Fig. 5c,d), which is in good agreement with the qRT-PCR analysis data. The densitometric analysis of the blots revealed that the decrease of MYC expression in cancer cells is in the range of $85 \pm 3\%$ and $52 \pm 2\%$ at 5.0 μM

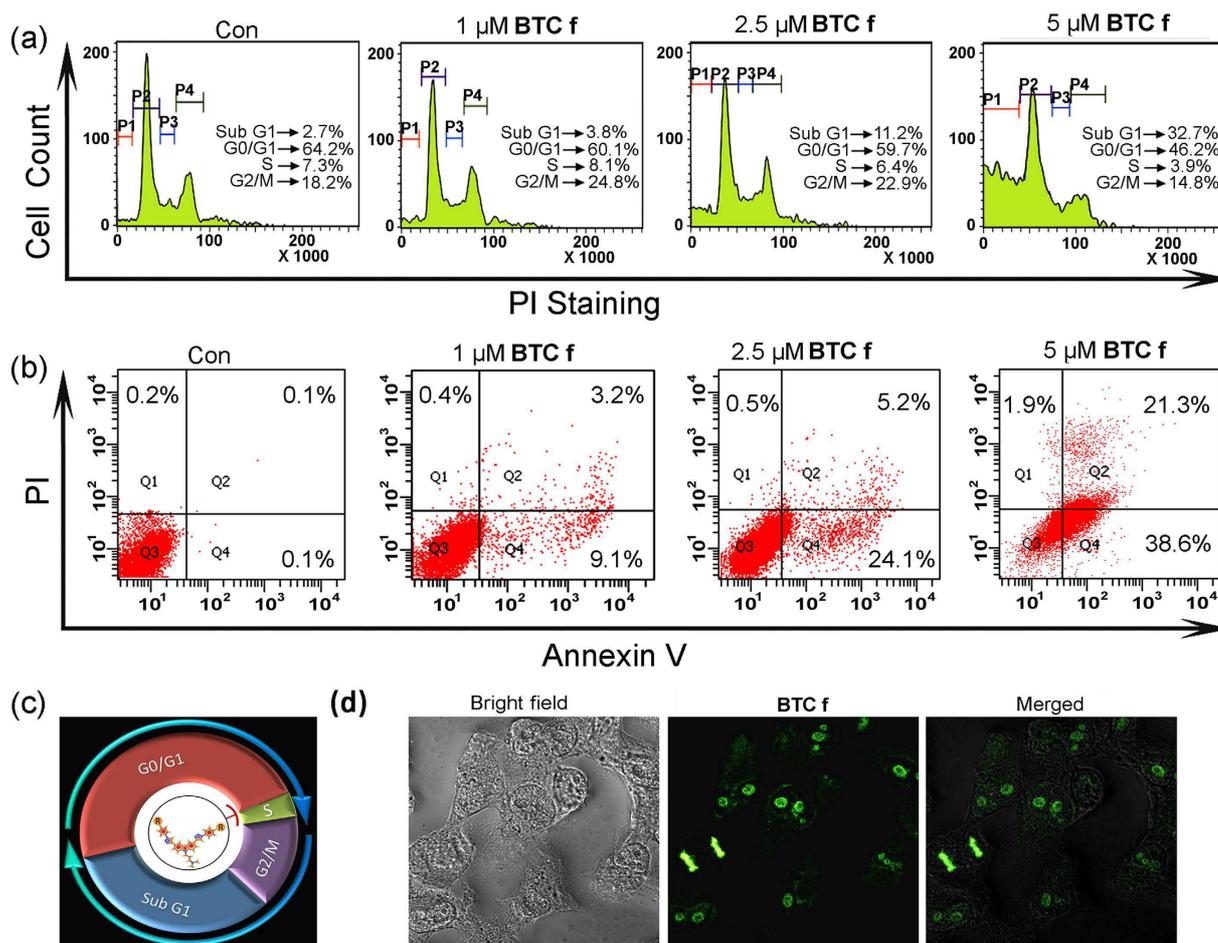


Figure 6. (a) Flow cytometric analysis of cell cycle parameters after incubation of **BTC f** (1.0–5.0 μM) in HepG2 cancer cells. P1, P2, P3 and P4 represent cell population at Sub G1, G0/G1, S and G2/M respectively. (b) Flow cytometric analysis of the mode of cancer cell death after treatment with **BTC f** (1.0–5.0 μM) in HepG2 cancer cells; Lower left (Q3), lower right (Q4), upper right (Q2) and upper left (Q1) quadrants indicate healthy cells, early, late apoptotic and necrotic cells, respectively. (c) Schematic representation of cell cycle arrest at Sub G1 phase by **BTC f**. (d) Confocal Microscope images (400 X magnification) showing localization of **BTC f**. All results are representative of three independent experiments with similar results.

concentration of **BTC f** and **BTC c**, respectively. The concentration required for 50% inhibition in *c-MYC* expression (IC_{50}^{MYC}) is approximately 2 μM for **BTC f** and 4.6 μM for **BTC c**. Almost negligible reduction in *GAPDH* expression was observed in both treated and untreated cells. Together these results suggest that the **BTC** ligands can reduce the expression of *MYC* protein in HepG2 cancer cells at mRNA and protein levels (Fig. 5a).

BTC f inhibits cell cycle progression and induces apoptosis. To determine whether **BTC f** mediated inhibition of HepG2 cell proliferation was associated with cell cycle arrest, we performed cell cycle analysis by propidiumiodide (PI) staining using flow cytometer. Flow cytometry analysis of **BTC f** (1.0–5.0 μM) treated HepG2 cells showed an increase in the SubG1 population (2.7% to 32.7%) with the increasing concentration of **BTC f** (Fig. 6a). Interestingly, at lower concentrations of **BTC f**, the cell population in G2 phase was increased (18.2% to 24.8%), and as the concentration of **BTC f** is increased, both S and G2 phase populations were decreased (7.3% to 3.9% and 18.2% to 14.8%, respectively) with a subsequent increase in the population of SubG1 phase.

To gain insight into the mechanism by which **BTC f** induced cell death of HepG2 cells, flow cytometry was employed to investigate the mode of cell death using Annexin-V and PI dual staining assay (Fig. 6b). HepG2 cells were incubated with **BTC f** (1.0–5.0 μM) for 24h and the untreated cells were used as control. The flow cytometry analysis revealed that the apoptotic cell population was increased significantly (0.2% to 59.9%) in a dose-dependent manner. However, the extent of necrotic death was only 1.9% under the experimental conditions. This is in well agreement with the reduced activity of *c-MYC* in liver carcinoma cells after treatment with **BTC f**.

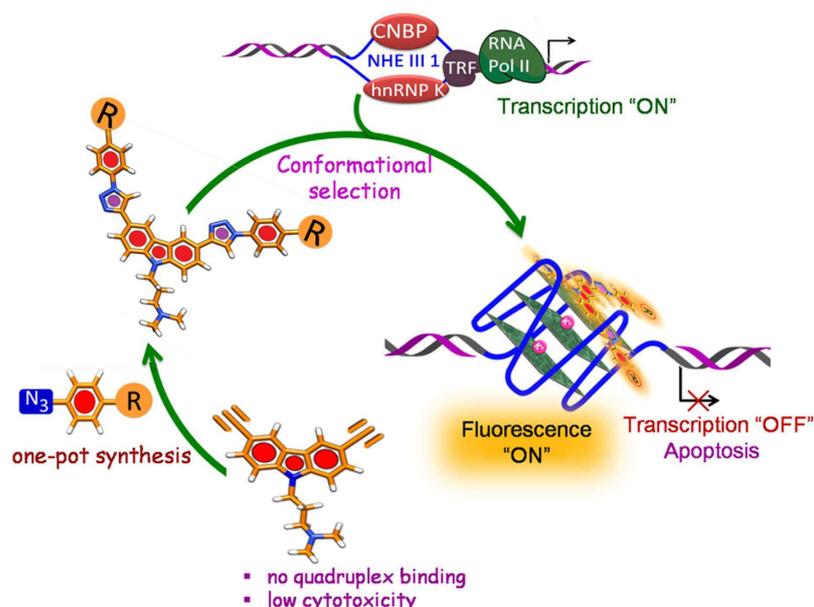


Figure 7. Bis-triazolylcarbazole derivatives stabilize *c-MYC* G-quadruplex and inhibit *c-MYC* transcription.

BTC f as a fluorescent probe for live-cell nucleus imaging. Confocal laser scanning microscopy (CLSM) was employed to examine the cellular localization of **BTC f** in living cells. HepG2 cells were treated with **BTC f** (5.0 μM) for 4 h and the CLSM images were taken, which clearly showed that the **BTC f** selectively stains the nucleus compared to the cytoplasm. The merged image (Fig. 6d) establishes the localization of **BTC f** within the nucleus. These data suggest that **BTC f** is cell-permeable and binds to the cellular DNA. As **BTC f** shows higher affinity towards the G-quadruplex over duplex DNA, its localization in the nucleus area indicates that **BTC f** may induce apoptosis by stabilizing the G-quadruplex in HepG2 cells.

Discussion

A modular access to bis-triazolyl carbazole derivatives has been devised, where a simple carbazole dialkyne precursor **5** was prepared from the commercially available carbazole. The carbazole dialkyne **5** was treated with a variety of azides **6a-g** using Cu(I) catalyzed azide-alkyne cycloaddition to prepare the corresponding bis-triazolylcarbazole derivatives **BTC a-g** in high yields (Fig. 1). FRET melting analysis (Fig. 2) revealed that the carbazole dialkyne **5** does not change the melting temperature of any of the three investigated promoter quadruplexes (*c-MYC*, *c-KIT1* and *c-KIT2*) or the duplex DNA, while the bis-triazolyl carbazole derivative **BTC f** with two carboxamide side chains exhibits high stabilization potential for the quadruplexes over the duplex DNA (Fig. 2 and Fig. 7). The diamino bis-triazolylcarbazole **BTC c** displayed moderate stabilization for the G-quadruplex sequences. Competitive FRET-melting experiments (Fig. 2b) clearly showed that the presence of 100 mol equivalent excess *ds* DNA did not significantly interfere with the stabilization of quadruplexes induced by the ligand **BTC f**, indicating its high selectivity for the quadruplexes over duplex DNA. FRET melting experiments at various concentrations of **BTC f** revealed that higher concentrations (5–7 fold) of **BTC f** are required to achieve the maximum stabilization potential (ΔT_m) for *c-KIT1* and *c-KIT2* quadruplexes compared to the *c-MYC* quadruplex. These results indicated that **BTC f** shows a preference for the *c-MYC* over the *c-KIT1* and *c-KIT2* quadruplexes. **BTC f** can also stabilize the quadruplexes to attain the maximum stabilization potential in the absence of K^+ ion. Despite showing high stabilization potential for the quadruplexes, **BTC f** did not exhibit any detectable stabilisation for the *ds* DNA.

Fluorescence studies showed that **BTC f** is a microenvironment-sensitive fluorescence "turn-on" sensor that selectively detects the *c-MYC* quadruplex over the duplex DNA with an enhancement in the fluorescence intensity with a blue shift (Fig. 3). However, the carbazole derivative **BTC c** in the presence of *c-MYC* neither triggered a significant fluorescence "turn-on" response nor a blue shift. The difference in fluorescence properties between **BTC c** and **BTC f** may be attributed to the lack of extended π -electron delocalization in **BTC c**. In agreement with the FRET melting results, the fluorescence spectroscopy studies showed high specificity of **BTC f** for the quadruplexes over duplex DNA, and in particular, **BTC f** showed a 5-fold preference for *c-MYC* ($K_D = 300 \text{ nM}$) over *c-KIT1* and *c-KIT2* promoter quadruplexes. Moreover, the confocal laser scanning microscopy (CLSM) shows that **BTC f** can be successfully applied to visualize the live-cell nucleus of HepG2 cells with high selectivity (Fig. 6).

NMR and Circular Dichroism (CD) spectroscopic analyses were employed to gain insight into the structural basis for the recognition of **BTC c** and **BTC f** to *c-MYC* quadruplex. CD binding titrations revealed that the binding of **BTC f** to *c-MYC* gives rise to a CD signal (ICD peak) from the bound ligand (Supplementary Fig. S5a-b). No such ICD bands are observed upon interaction of **BTC f** with the *c-KIT1* and *c-KIT2* quadruplexes, which indicates that **BTC f** binds to the *c-MYC* in an asymmetric fashion. It has been reported that multiple conformations of 3'- and 5'-flanking capping structures of *c-MYC* quadruplex co-exist at equilibrium¹³. NMR analyses (Fig. 4) suggested that **BTC f** selects and binds specifically to one of these minor conformations of the *c-MYC* via conformational selection. The [**BTC f**]:[*c-MYC*] interaction process is similar with or without K⁺. The binding of **BTC f** could also induce a conformational change in the binding-competent minor conformation of the *c-MYC*, which may be useful to specifically alter the biological function of the *c-MYC*.

Among the synthesized carbazole analogues, compound **BTC f** was found to be the most potent molecule that inhibited the growth of human hepatocellular liver carcinoma HepG2 cells (which highly expresses MYC) at a low micromolar concentration (IC₅₀ = 4.3 ± 0.69 μM) without affecting the normal mouse myoblast C2C12 cells. The less potent quadruplex stabilizer **BTC c** also showed good antiproliferative activity with an IC₅₀ value of 11.38 ± 1.54 μM in HepG2 cells. It is worth mentioning that the starting carbazoledialkyne **5** neither stabilized the G-quadruplex DNA (FRET melting data) nor effectively inhibited the growth of cancer cells (Fig. 7). Since **BTC f** exhibits high specificity for the *c-MYC* quadruplex and the *c-MYC* gene is overexpressed in liver carcinoma HepG2 cells, the effect of **BTC f** and the less potent **BTC c** on transcriptional regulation *c-MYC* gene was studied in HepG2 cells (Fig. 5). **BTC f** was able to inhibit *c-MYC* expression in both transcriptional and translational level as suggested by qRT-PCR and Western blotting analysis (Fig. 5). Analysis of the cell cycle data revealed that cells were unable to transit from G1 phase to S phase with increasing concentration of **BTC f**. The increase in concentration of **BTC f** triggered cell cycle arrest in SubG1 phase and subsequent apoptosis in cancer cells, probably by down-regulating *c-MYC* gene expression^{41–43} as suggested by qRT-PCR and Western blot analysis (Figs 5 and 6). Our results collectively suggest that the reduction in *c-MYC* expression is probably due to the binding of **BTC f** to the *c-MYC* promoter-quadruplex. However, the exact molecular mechanism of **BTC f** mediated *c-MYC* down-regulation⁴⁴ is currently under investigation.

This work highlights that a simple synthetic protocol can be devised to synthesize fluorescent bis-triazolylcarbazole derivatives, which can effectively inhibit *c-MYC* transcriptional activity (Fig. 7). The carbazole derivative with carboxamide side chains shows pronounced environment-sensitive fluorescence and selectively detects the *c-MYC* G-quadruplex over the duplex DNA via “turn-on” fluorescence. The ligand binds to the *c-MYC* with nM binding affinity and shows a five-fold preference for the *c-MYC* over the *c-KIT* promoter quadruplexes. The ligand can bind and stabilize the *c-MYC* quadruplex in the presence and absence of added K⁺. It is intriguing that the ligand can be used to ‘programme’ the *c-MYC* to adopt a specific conformation, which is stable with or without K⁺. The specific nucleic acid structure may find applications in the fields of nanobiotechnology and biomedical technology. This small molecule probe is an attractive probe for bio-imaging as it can rapidly and selectively stain the nucleus in living cells. Further this small molecule can induce cell cycle arrest and promote cancer cell death by apoptosis. These results collectively suggest that the carbazole derivative is a potent anticancer agent and a viable lead for further development of anticancer drugs.

Synthetic Protocols. *Synthesis of 3-(3,6-diiodo-9H-carbazol-9-yl)-N,N-dimethylpropan-1-amine 3.* A mixture of carbazole **1** (15.0 g, 89.82 mmol), KI (19.36 g, 116.62 mmol), KIO₃ (19.20 g, 89.82 mmol) in acetic acid (100 mL) and deionized water (10 mL) was stirred at 80 °C for 48 h under N₂ atmosphere. After cooling to room temperature, the mixture was filtered, washed with deionized water and saturated Na₂CO₃ solution to afford **3**, 6-diiodocarbazole as a colorless solid (24.4 g, 65%). ¹H NMR (500 MHz, CDCl₃): 8.32 (s, 2H), 8.09 (s_{br}, 1H), 7.68 (d, 2H, J = 8.4 Hz), 7.21 (d, 2H, J = 8.4 Hz); ¹³C NMR (100 MHz, CDCl₃): 138.5, 134.8, 129.4, 124.6, 112.7, 82.4.

A mixture of **3**, 6-diiodocarbazole (500 mg, 1.193 mmol) and NaH (114.24 mg, 4.77 mmol) in 20 mL THF was stirred at room temperature for 2 h under nitrogen atmosphere. In another flask, 3-dimethylaminopropyl chloride hydrochloride **2** (435.33 mg, 3.58 mmol) and NaOH (1.0 g) were diluted in 5 mL water and cooled to room temperature, then the upper layer of this solution was added drop-wise to the above mixture. The reaction mixture was refluxed for an additional 24 h. After the removal of the solvent under reduced pressure, the residue was purified by column chromatography on silica gel using DCM/methanol (20:1) as eluent. Re-crystallization from ethanol gave the compound **3** (529 mg, 88%) as a colorless solid, mp 129–131 °C. ¹H NMR (500 MHz, CDCl₃): 8.32 (s, 2H), 7.70 (d, 2H, J = 8.4 Hz), 7.25 (d, 2H, J = 9.2 Hz), 4.32 (t, 2H, J = 6.7 Hz), 2.20 (s, 8H), 1.96 (t, 2H, J = 6.7 Hz); ¹³C NMR (100 MHz, CDCl₃): 139.6, 134.5, 129.3, 124.0, 111.0, 81.8, 56.1, 45.2, 40.6, 26.5. HRMS (ESI) calculated for [C₁₇H₁₈I₂N₂]: 504.9632, Found 504.9635.

Synthesis of 3-(3,6-diethynyl-9H-carbazol-9-yl)-N,N-dimethylpropan-1-amine 5. To a solution of **2** (300 mg, 0.595 mmol) in Et₃N (5 mL) added PdCl₂(PPh₃)₂ (35.1 mg, 0.05 mmol) and CuI (19.04 mg, 0.1 mmol) at room temperature. The mixture was stirred for 30 min and then 2-methylbut-3-yn-2-ol **4** (0.182 mL, 1.785 mmol) was added drop-wise. The resulting mixture was stirred under an argon

atmosphere for 12 h, concentrated, washed with brine and dried over anhydrous Na_2SO_4 . The crude product was purified by column chromatography. Then the resulting crude alcohol was refluxed with 5 equiv. KOH in toluene under an argon atmosphere for 12 h. The reaction mixture was concentrated, washed with brine and dried over anhydrous Na_2SO_4 . The crude product was purified by column chromatography to give the dialkyne **5** (161 mg, 90% yield) as a yellow liquid. ^1H NMR (500 MHz, CDCl_3): 8.22 (s, 2H), 7.60 (d, 2H, $J = 8.4$ Hz), 7.41 (d, 2H, $J = 9.3$ Hz), 4.38 (t, 2H, $J = 6.7$ Hz), 3.08 (s, 2H), 2.22 (s, 8H), 1.99 (t, 2H, $J = 6.7$ Hz); ^{13}C NMR (100 MHz, CDCl_3): 141.0, 130.4, 124.9, 122.5, 113.0, 109.4, 85.0, 75.7, 56.4, 45.6, 41.0, 27.0; IR (KBr, cm^{-1}): 3284, 3269, 2925, 2817, 2767, 2100, 1627, 1595, 1481, 1344, 1286, 1249, 1211, 1134; HRMS (ESI) calculated for $[\text{C}_{21}\text{H}_{21}\text{N}_2]$: 301.1699, Found 301.1696.

General procedure for the synthesis of bis-triazole derivatives BTC a-g using click chemistry (GP1). Dialkyne **5** (50 mg, 0.166 mmol) was dissolved in a 1:2 mixture of *t*-BuOH/ H_2O (4 mL). Copper(II) sulphate pentahydrate (4.14 mg, 0.0166 mmol) and sodium ascorbate (3.2 mg, 0.0166 mmol) were added and the solution was stirred for 10 min. The corresponding azide **6** (2.5×0.166 mol equiv.) was added and the mixture was then heated for 4 h at 70°C under microwave irradiation. After cooling to room temperature, the reaction mixture was concentrated. The crude product was purified by flash column chromatography (from CH_2Cl_2 (100%) to $\text{CH}_2\text{Cl}_2/\text{MeOH}$ (10:1) to $\text{CH}_2\text{Cl}_2/\text{MeOH}/\text{NH}_4\text{OH}$ (10:1:0.5)) to give the corresponding bis-triazole derivatives **BTC a-g**.

Synthesis of BTC a. Following the **GP-1**, the reaction of the dialkyne **5** with azide **6a** (49 mg) afforded **BTC a** (55 mg, 61%) as a yellow viscous liquid. ^1H NMR (500 MHz, $\text{DMSO}-d_6$): 9.35 (s, 2H), 8.84 (s, 2H), 8.10 (d, 2H, $J = 8.8$ Hz), 8.00 (d, 4H, $J = 7.8$ Hz), 7.78 (d, 2H, $J = 8.8$ Hz), 7.66 (t, 4H, $J = 6.4$ Hz), 7.53 (t, 2H, $J = 6.8$ Hz), 4.50 (t, 2H, $J = 7.3$ Hz), 2.26 (t, 2H, $J = 5.8$ Hz), 2.17 (s, 6H), 1.98 (t, 2H, $J = 6.8$ Hz); ^{13}C NMR: (125 MHz, $\text{DMSO}-d_6$): 148.3, 140.4, 136.8, 129.9, 128.5, 123.8, 122.4, 121.5, 119.8, 118.6, 117.3, 110.0, 55.8, 44.9, 40.2 (merged with $\text{DMSO}-d_6$), 22.0; IR (KBr, cm^{-1}): 3342, 3303, 2960, 2852, 1641, 1629, 1591, 1514, 1481, 1197, 1033, 1001; HRMS (ESI) calculated for $[\text{C}_{33}\text{H}_{31}\text{N}_8]$: 539.2672, Found 539.2664.

Synthesis of BTC b. Following the **GP-1**, the reaction of the dialkyne **5** with azide **6b** (56 mg) afforded **BTC b** (70 mg, 74%) as a deep brown solid, mp $> 220^\circ\text{C}$; ^1H NMR (500 MHz, $\text{DMSO}-d_6$): 10.02 (s, 2H), 9.16 (s, 2H), 8.81 (s, 2H), 8.07 (d, 2H, $J = 8.4$ Hz), 7.74–7.77 (m, 6H), 7.00 (d, 4H, $J = 8.4$ Hz), 4.49 (t, 2H, $J = 6.7$ Hz), 2.32 (t, 2H, $J = 6.7$ Hz), 2.20 (s, 6H), 1.99 (t, 2H, $J = 6.7$ Hz); ^{13}C NMR: (125 MHz, $\text{DMSO}-d_6$): 157.7, 148.0, 140.3, 128.9, 123.7, 122.4, 121.7, 118.5, 117.2, 116.1, 109.9, 55.8, 44.7, 40.8 (merged with $\text{DMSO}-d_6$), 26.2; IR (KBr, cm^{-1}): 3126, 3026, 2952, 2831, 2790, 1604, 1521, 1481, 1465, 1452, 1282, 1220, 1164, 1058, 995; HRMS (ESI) calculated for $[\text{C}_{33}\text{H}_{31}\text{N}_8\text{O}_2^+]$: 571.2564, Found 571.2567.

Synthesis of BTC c. Following the **GP-1**, the reaction of the dialkyne **5** with azide **6c** (56 mg) afforded **BTC c** (65 mg, 68.5%), a deep yellow solid, mp $110\text{--}112^\circ\text{C}$; ^1H NMR (500 MHz, $\text{DMSO}-d_6$): 9.04 (s, 2H), 8.79 (s, 2H), 8.06 (dd, 2H, $J = 8.5, 1.2$ Hz), 7.73 (d, 2H, $J = 8.2$ Hz), 7.58 (d, 4H, $J = 8.8$ Hz), 6.74 (d, 4H, $J = 8.8$ Hz), 5.52 (s, 4H), 4.48 (t, 2H, $J = 6.3$ Hz), 2.22 (t, 2H, $J = 6.9$ Hz), 2.14 (s, 6H), 1.96 (t, 2H, $J = 6.9$ Hz); ^{13}C NMR: (125 MHz, $\text{DMSO}-d_6$): 149.3, 147.8, 140.3, 126.2, 123.7, 122.5, 121.9, 121.4, 118.2, 117.2, 113.9, 109.9, 55.9, 44.8, 40.8 (merged with $\text{DMSO}-d_6$), 26.4; IR (KBr, cm^{-1}): 3338, 3134, 2702, 2441, 1623, 1521, 1477, 1220, 1043; HRMS (ESI) calculated for $[\text{C}_{33}\text{H}_{32}\text{N}_{10}]$: 569.2884, Found 569.2886.

Synthesis of BTC d. Following the **GP-1**, the reaction of the dialkyne **5** with azide **6d** (61 mg) afforded **BTC d** (65 mg, 66%) as an orange solid, mp $> 220^\circ\text{C}$; ^1H NMR (500 MHz, $\text{DMSO}-d_6$): 10.10 (s, 2H), 9.52 (s, 2H), 8.85 (s, 2H), 8.26 (d, 4H, $J = 8.5$ Hz), 8.19 (d, 4H, $J = 8.6$ Hz), 8.10 (d, 2H, $J = 8.6$ Hz), 7.80 (d, 2H, $J = 8.6$ Hz), 4.51 (t, 2H, $J = 6.1$ Hz), 2.36 (s, 2H), 2.23 (s, 6H), 2.01 (t, 2H, $J = 6.1$ Hz); ^{13}C NMR: (125 MHz, $\text{DMSO}-d_6$): 192.1, 148.7, 140.6, 135.6, 131.3, 123.9, 122.5, 121.2, 119.9, 118.7, 117.4, 112.9, 110.1, 55.7, 43.6, 40.8 (merged with $\text{DMSO}-d_6$), 26.1; IR (KBr, cm^{-1}): 2954, 2852, 1697, 1602, 1481, 1406, 1307, 1209, 1163, 1033; HRMS (ESI) calculated for $[\text{C}_{35}\text{H}_{31}\text{N}_8\text{O}_2]$: 595.2564, Found 595.3433.

Synthesis of BTC e. Following the **GP-1**, the reaction of the dialkyne **5** with azide **6e** (68 mg) afforded **BTC e** (70 mg, 67%) as a light yellow solid, mp $> 220^\circ\text{C}$; ^1H NMR (500 MHz, $\text{DMSO}-d_6$): 10.71 (s, 2H), 9.49 (s, 2H), 8.85 (s, 2H), 8.21–8.10 (m, 10H), 7.86 (d, 2H, $J = 8.4$ Hz), 4.59 (s, 2H), 3.20 (s, 2H), 2.73 (s, 6H), 2.50 (s, 2H); ^{13}C NMR: (125 MHz, $\text{DMSO}-d_6$): 166.4, 148.6, 140.3, 139.6, 131.1, 130.5, 123.9, 122.6, 121.6, 119.5, 118.7, 117.4, 110.2, 54.2, 48.5, 39.7 (merged with $\text{DMSO}-d_6$), 23.7; IR (KBr, cm^{-1}): 3406, 3126, 2628, 2113, 1703, 1606, 1481, 1382, 1224, 1039; HRMS (ESI) calculated for $[\text{C}_{35}\text{H}_{31}\text{N}_8\text{O}_4]$: 627.2463, Found 627.2465.

Synthesis of BTC f. Following the **GP-1**, the reaction of the dialkyne **5** with azide **6f** (102 mg) afforded **BTC f** (96 mg, 73%) as a yellow solid, mp $69\text{--}71^\circ\text{C}$; ^1H NMR (500 MHz, $\text{DMSO}-d_6$): 9.43 (s, 2H), 8.85 (s, 2H), 8.70 (t, 2H, $J = 5.9$ Hz), 8.11 (s, 8H), 8.10 (s, 2H), 7.79 (d, 2H, $J = 8.4$ Hz), 4.51 (t, 2H, $J = 5.9$ Hz), 3.32 (4H, merged with water peak), 2.29 (t, 4H, $J = 6.7$ Hz), 2.22 (t, 2H, $J = 6.7$ Hz), 2.16 (s, 18H), 1.98 (t, 2H, $J = 6.7$ Hz), 1.67–1.72 (m, 4H); ^{13}C NMR: (125 MHz, $\text{DMSO}-d_6$): 165.1, 148.6, 140.6, 138.4, 134.4, 129.0, 122.5, 121.4, 119.4, 118.6, 117.4, 110.1, 56.9, 56.0, 43.6, 40.9 (merged with $\text{DMSO}-d_6$), 38.9, 27.0,

26.5; IR (KBr, cm^{-1}): 3380, 3062, 2657, 2432, 1645, 1608, 1548, 1508, 1481, 1296, 1224, 1037; HRMS (ESI) calculated for $[\text{C}_{45}\text{H}_{55}\text{N}_{12}\text{O}_2]$: 795.4565, Found 795.4778.

Synthesis of BTC g. Following the **GP-1**, the reaction of dialkyne **5** with azide **6g** (36 mg) afforded **BTC g** (60 mg, 76%) as a brown liquid; ^1H NMR (500 MHz, DMSO-d_6): 8.72 (s, 2H), 8.56 (s, 2H), 7.98 (dd, 2H, $J = 8.5, 1.5$ Hz), 7.68 (d, 2H, $J = 8.5$ Hz), 5.16 (s, 2H), 4.49–4.43 (m, 6H), 3.87 (t, 4H, $J = 4.9$ Hz), 2.20–2.10 (m, 8H), 1.94 (t, 2H, $J = 6.7$ Hz); ^{13}C NMR: (125 MHz, DMSO-d_6): 147.1, 140.1, 123.5, 122.4, 122.2, 120.9, 117.1, 109.8, 59.9, 55.9, 52.4, 45.1, 40.8 (merged with DMSO-d_6), 27.8; IR (KBr, cm^{-1}): 3271, 2667, 1600, 1481, 1465, 1244, 1135, 887; HRMS (ESI) calculated for $[\text{C}_{25}\text{H}_{31}\text{N}_8\text{O}_2]$: 475.2564, Found 475.2576.

Online Methods. *FRET melting experiments.* Stock solutions having 1 μM concentration of each carbazole compound was prepared in MQ water (**BTC f**) and DMSO (compound **5**, **BTC a–e** and **BTC g**). Four dual fluorescently labeled DNA oligonucleotide sequences were diluted in 50 mM potassium cacodylate buffer, pH 7.4.

c-MYC: 5'-FAM-d(TGAG₃TG₃TAG₃TG₃TA₂)-TAMRA-3'

c-KIT1: 5'-FAM-d(GGGAGGGCGCTGGGAGGGAGGG)-TAMRA-3',

c-KIT2: 5'-FAM-d(GGGCGGGCGCGAGGGAGGGG)-TAMRA-3' and

ds DNA: 5'-FAM-d(CCAGTTCGTAGTAACCC)-3'/3'-TAMRA (GGTCAAGCATCATTGGG)-5'

The donor fluorophore was 6-carboxyfluorescein, FAM, and the acceptor fluorophore was 6-carboxy tetramethylrhodamine, TAMRA. Dual-labeled DNA was annealed at a concentration of 200 nM by heating at 95 °C for 5 min followed by cooling to room temperature. The 96-well plates were prepared by aliquoting 50 μL of the annealed DNA into each well, followed by 50 μL of the carbazole compounds.

For competition experiments, duplex competitors were added to 200 nM quadruplex sequences at final concentration of 200 nM, 1.0 μM , 2.0 μM and 20 μM . The concentration of **BTC f** was kept at 1.0 μM . For FRET titration experiments different concentration of **BTC f** (0.05, 0.1, 0.2, 0.4, 0.6, 0.8, 1.0 μM) were added to 100 nM of all four DNA sequences. Measurements were made in triplicate with an excitation wavelength of 483 nm and a detection wavelength of 533 nm using a LightCycler® 480-II System RT-PCR machine (Roche). Final analysis of the data was carried out using Origin Pro 8 data analysis.

Fluorimetric titration. Fluorescence emission spectra were recorded with successive addition of the quadruplex solution into the ligand solution. The fluorescence spectra were recorded on a Horiba JobinYvonFluoromax 3 instrument at 25 °C in a thermostated cell holder using quartz cuvette with a 1 cm path-length. The spectra were taken using filtered and degassed buffer (100 mM Tris-HCl containing 100 mM KCl) and other solvents like ethyl acetate, DMF, DMSO, ethanol, methanol and water. In this experiment, 1.0 μM **BTC f** was titrated with aliquots of different preannealed DNA sequences:

c-MYC: 5'-d(TGAG₃TG₃TAG₃TG₃TA₂)-3'

c-KIT1: 5'-d(GGGAGGGCGCTGGGAGGGAGGG)-3',

c-KIT2: 5'-d(GGGCGGGCGCGAGGGAGGGG)-3' and

ds DNA: 5'-d(CCAGTTCGTAGTAACCC)-3'/3'-(GGTCAAGCATCATTGGG)-5'

The recorded spectral data was used to determine the dissociation constant of the ligands for quadruplexes using the Hill-1 formula (eqn. 1):

$$F = F_0 + \frac{(F_{max} - F_0)[DNA]}{K_D + [DNA]} \quad (1)$$

F is the fluorescence intensity, F_{max} is the maximum fluorescence intensity, F_0 is the fluorescence intensity in the absence of DNA and K_D is the dissociation constant.

CD spectroscopy. CD spectra were recorded on a JASCO J-815 spectrophotometer by using a 1 mm path length quartz cuvette. Aliquots of **BTC f** were added stepwise to pre-annealed *c-MYC* (TGAG₃TG₃TAG₃TG₃TA₂), *c-KIT1* (GGGAGGGCGCTGGGAGGGAGGG) and *c-KIT2* (GGGCGGGCGCGAGGGAGGGG) quadruplex sequences in Tris•HCl (100 mM) buffer at pH 7.4 containing KCl (100 mM). The CD spectra were also recorded upon incremental addition of **BTC f** to the *c-MYC* quadruplex sequence in Tris•HCl buffer (100 mM), at pH 7.4 in the absence of KCl. The DNA concentrations used were 10 μM . The CD spectra represent an average of three scans and were smoothed and zero corrected. Final analysis and manipulation of the data was carried out by using Origin 8.0.

NMR spectroscopy. NMR experiments were performed using intramolecular G-quadruplex formed by *c-MYC* purchased by Eurofins MWG Operon (Ebersberg, Germany) as HPSF® (High Purity Salt Free) purified oligos and further purified *via* HPLC. Sequence and numbering of the oligonucleotide used for NMR studies is reported in Fig. 4e, while folding topology of the major conformation of *c-MYC* determined by NMR by Ambrus *et al.* is shown in Fig. 4d)¹³. The proton resonance assignment of the *c-MYC* DNA in 90% H₂O/10% D₂O was performed on the basis of the assignment reported in literature.¹³ NMR samples were referenced with 2,2-dimethyl-2-silapentane-5-sulphonate (DSS) and prepared in buffer containing 25 mM tris-HCl at pH 7.4 with or without additional 100 mM KCl, in 90% H₂O/10% D₂O.

^1H NMR spectra were recorded on DNA and DNA-ligand complexes with gradient-assisted excitation sculpting⁴⁵ or jump-return-Echo⁴⁶ for water suppression.

Cell viability analysis using MTT assay. The human hepatocellular carcinoma cells (HepG2), human breast cancer cells (MCF 7), human colon carcinoma cells (HCT 116) and mouse normal myoblast cells (C2C12) were cultured in DMEM containing high glucose (5.5 mM) supplemented with 10% FBS at pH 7.4. Cells were maintained in tissue culture plates containing 4×10^5 cells/well at 37 °C in an atmosphere of 5% carbon dioxide (CO_2)/95% air. The MTT cell proliferation assay determines the ability of living cells to reduce the yellow tetrazole, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) to purple formazan crystals by mitochondrial enzymes. For the MTT assay, HepG2 cancer cells were treated with various concentrations of carbazole derivatives **5** and **BTC a–g** for 24 h. Following incubation with each compound for 24 h, 20 μL of MTT was added (at a concentration of solution 5 mg/mL in phosphate-buffered saline, pH 7.4.) to each well. After incubation for 4 h at 37 °C, the culture medium was removed, and the formazan crystals were dissolved in 200 μL DMSO. Absorbance (A) of formazan dye was measured at 570 nm using a micro plate reader. The background absorbance was determined at 690 nm and subtracted from the 570 nm measurement. The percentage of viable cells was determined by the equation (2):

$$\text{Viable cells (\%)} = \frac{A \text{ of treated cells}}{A \text{ of untreated cells}} \times 100 \quad (2)$$

Following the similar protocol, MCF 7, HCT 116 and C2C12 were treated with various concentrations of **BTC f** and the percentage of viable cells was determined.

qRT-PCR analysis. To evaluate the role of **BTC** at the transcriptional level of *c-MYC*, qRT-PCR was performed. Cancer cells were incubated with various concentrations (1.0, 2.5 and 5.0 μM) of **BTC c** and **BTC f** for 24 h at 37 °C in humidified 5% CO_2 incubator. Total RNA was prepared from compound treated and untreated HepG2 cells using the Trizol kit according to the manufacturer's protocol (Invitrogen Corporation). cDNA library was prepared by High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). Master mix (2X) was prepared by 2 μL 10X RT buffer, 0.8 μL 25 \times 100 mM dNTP mix, 2 μL 10X RT random primer, 1 μL of reverse transcriptase, 1 μL of RNase inhibitor and 3.2 μL nuclease free water. 10 μL 2X master mix was added to 10 μL sample and sealed in 96 well plates. Reverse transcriptase reaction was performed on a Light Cycler 480 II (Roche). The thermal cycling condition was programmed as 10 min at 25 °C, 120 min at 37 °C and 5 min at 85 °C for one single cycle. qRT-PCR was performed on a Light Cycler 480 II (Roche) with SYBR green JumpStart Taq ReadyMix (Sigma, Saint Louis, USA) reagent using the cDNA library as template. The primers used for the qRT-PCR analyses had the following sequences^{47,48}:

c-MYC (forward): 5'-CTGCGACGAGGAGGAGGACT-3'

c-MYC (reverse): 5'-GGCAGCAGCTCGAATTTCTT-3'

GAPDH (forward): 5'-GACGGCCGCATCTTCTTGT-3'

GAPDH (reverse): 5'-CACACCGACCTTACCATT-3'

The PCR mixture (25 μL) contained 15 pmol of each primer, 7 μL of water, 5 μL of cDNA, 12.5 μL 2X JumpStart Taq ReadyMix. The samples were placed in 96-well plates (Roche), and PCR amplification was performed using Light Cycler 480 II real-time PCR detection system (Roche). The thermal cycling conditions were 2 min at 94 °C and then 40 cycles of 15 s at 94 °C, followed 60 s at 60 °C. We used the comparative cycle threshold method (C_T method) for relative quantification of gene expression⁴⁹. The C_T for the target and the C_T for the internal control (GAPDH) were determined for **BTC c** or **BTC f** or untreated (control) samples. Finally, the arithmetic calibrator ($2^{-\Delta\Delta C_T}$) was used to calculate the relative mRNA level expression of *c-MYC*. Difference in *c-MYC* expression was expressed as fold changes.

Western blot analysis. Hepatocellular carcinoma HepG2 cells were treated with different concentration (1.0, 2.5 and 5.0 μM) of **BTC c** and **BTC f** for 24 h at 37 °C in humidified CO_2 incubator. After the incubation period, cells were washed once with PBS (pH 7.4) and lysed with cold cell lysis buffer (20 mM Tris, 100 mM NaCl, 1 mM EDTA in 0.5% Triton X-100). Cell lysate were collected from the treated and untreated cells, and the total protein content was estimated by Lowry method⁵⁰. Equal amount of proteins (60 μg) from the cell lysate were separated by 10% SDS-PAGE and transferred to nitrocellulose membranes. The membranes were blocked, washed and probed using antibodies directed against *c-MYC* and GAPDH (as loading control) for overnight at room temperature. The blots were washed and immunoreactive bands were incubated with a 1:2000 dilution of ALKP conjugated secondary antibody for 2 h at room temperature. Binding signals were visualized with NBD/BCIP substrate. Relative band intensities were determined by using ImageJ software.

Flow cytometric determination of the cell cycle histogram by PI staining. Cell cycle histogram was analysed using propidium iodide (PI) staining by Flowcytometry. HepG2 cells (1×10^6) per 60 mm petridish (~80% confluence) were treated with various concentration of **BTC f** (1.0, 2.5, 5.0 μM) for 24 h in fresh

growth medium. Cells were harvested by trypsinization, resuspended in PBS and fixed with 2 mL of ice-cold 70% ethanol for overnight at 4 °C. The pellets were collected by centrifugation and resuspended in PBS solution, containing 10 µg/mL PI (Sigma) and 10 µg/mL RNaseA (Sigma). After incubation for 30 minutes in the dark at 37 °C, cells were analyzed for DNA content using a FACS flow cytometer (BD Biosciences). Cell distribution among cell cycle phases and the percentage of apoptotic cells were evaluated using Cell-Quest Pro software (BD).

Flowcytometric assay of apoptosis. Annexin V-FITC and propidium iodide (PI) stains were used to determine the percentage of cells undergoing apoptosis and necrosis. An apoptosis assay was conducted using the protocol supplied by the manufacturer. Briefly, 1×10^6 HepG2 cells per 60 mm petridish (~90% confluence) were treated with different concentration of **BTC f** (1.0, 2.5, 5.0 µM) for 24 h in fresh growth medium. Cells were then harvested with trypsinization. After centrifugation at 700 rpm for 5 min at 4 °C, cell pellet was suspended in 500 µL $1 \times$ binding buffer and then treated with 5 µL Annexin V-FITC and 5 µL PI. After incubation for 5 min on ice, each sample was analysed immediately using fluorescence-activated cell sorting (FACS) analysis (BD Biosciences, Mountain View, CA, USA). Approximately 10,000 HepG2 cells were detected for each sample. Cytoqram analysis was done using the Cell Quest software.

Confocal microscopy. Cellular localization of **BTC f** was monitored by live cell imaging. HepG2 cells (1×10^3) were cultured on 35 mm diameter glass-bottomed cover slips for 24 h followed by incubation with **BTC f** (5.0 µM) for 4 h inside CO₂ (5%) incubator at 37 °C. After incubation, cells were washed with PBS three times to remove the excess ligand and bathed in DMEM (2 mL) before imaging. Localization of **BTC f** was viewed under confocal microscope (Zeiss LSM 510 laser scanning microscope, Standort Göttingen, Germany). At least 5 fields per slide and three independent sets were examined.

Statistical analysis. All data were given as mean \pm S.D. Differences between two groups were compared by unpaired Student's t-test. For multi-group comparisons, analysis of variance was determined by ANOVA. A value of $P < 0.05$ was considered statistically significant. The statistical analysis was done by using GraphPad Instant Software (Graph-Pad, La Jolla, CA, USA).

References

- González, V. & Hurley, L. H. The c-MYC NHE III1: Function and Regulation. *Annu. Rev. Pharmacol. Toxicol.* **50**, 111–129 (2010).
- Kawate, S., Fukusato, T., Ohwada, S., Watanuki, A. & Morishita, Y. Amplification of c-myc in hepatocellular carcinoma: correlation with clinicopathologic features, proliferative activity and p53 overexpression. *Oncology* **57**, 157–163 (1999).
- Wu, K.-J. *et al.* Direct activation of TERT transcription by c-MYC. *Nat Genet* **21**, 220–224 (1999).
- Shachaf, C. M. *et al.* MYC inactivation uncovers pluripotent differentiation and tumour dormancy in hepatocellular cancer. *Nature* **431**, 1112–1117 (2004).
- Cooney, M., Czernuszewicz, G., Postel, E. H., Flint, S. J. & Hogan, M. E. Site-specific oligonucleotide binding represses transcription of the human c-myc gene *in vitro*. *Science* **214**, 456–459 (1988).
- Brooks, T. A. & Hurley, L. H. The role of supercoiling in transcriptional control of MYC and its importance in molecular therapeutics. *Nat. Rev. Cancer* **9**, 849–861 (2009).
- Seenisamy, J. *et al.* The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4. *J Am. Chem. Soc.* **126**, 8702–8709 (2004).
- Hatzakis, E., Okamoto, K. & Yang, D. Thermodynamic Stability and Folding Kinetics of the Major G-Quadruplex and Its Loop Isomers Formed in the Nuclease Hypersensitive Element in the Human c-Myc Promoter: Effect of Loops and Flanking Segments on the Stability of Parallel-Stranded Intramolecular G-Quadruplexes. *Biochemistry* **49**, 9152–9160 (2010).
- Mathad, R. I., Hatzakis, E., Dai, J. & Yang, D. 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**, 9023–9033 (2011).
- Sun, D. & Hurley, L. H. 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 (2009).
- Phan, A. T., Modi, Y. S. & Patel, D. J. Propeller-Type Parallel-Stranded G-Quadruplexes in the Human c-myc Promoter. *J. Am. Chem. Soc.* **126**, 8710–8716 (2004).
- Phan, A. T., Kuryavyi, V., Gaw, H. Y. & Patel, D. J. Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. *Nat Chem. Biol.* **1**, 167–173 (2005).
- Ambrus, A., Chen, D., Dai, J., Jones, R. A. & Yang, D. Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry* **44**, 2048–2058 (2005).
- Siddiqui-Jain, A., Grand, C. L., Bearss, D. J. & Hurley, L. H. 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 (2002).
- Collie, G. W. & Parkinson, G. N. The application of DNA and RNA G-quadruplexes to therapeutic medicines. *ChemSoc Rev.* **40**, 5867–5892 (2011).
- Balasubramanian, S., Hurley, L. H. & Neidle, S. Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nature Rev. Drug Discovery.* **10**, 261–275 (2011).
- Drygin, D. *et al.* Anticancer Activity of CX-3543: A Direct Inhibitor of rRNA Biogenesis. *Cancer Res.* **69**, 7653–7661 (2009).
- Ji, X. *et al.* The interaction of telomeric DNA and C-myc22 G-quadruplex with 11 natural alkaloids. *Nucleic Acid Ther.* **22**, 127–136 (2012).
- Pivetta, C. *et al.* Perylene side chains modulate G-quadruplex conformation in biologically relevant DNA sequences. *Bioorg Chem.* **16**, 9331–9339 (2008).
- Seenisamy, J. *et al.* Design and synthesis of an expanded porphyrin that has selectivity for the c-MYC G-quadruplex structure. *J. Am. Chem. Soc.* **127**, 2944–2959 (2005).
- Ou, T. M. *et al.* Stabilization of G-quadruplex DNA and down-regulation of oncogene c-myc by quindoline derivatives. *J Med Chem* **50**, 1465–1474 (2007).

22. Dash, J., Shirude, P. S., Hsu, S.-T. D. & Balasubramanian, S. Diarylethynyl Amides That Recognize the Parallel Conformation of Genomic Promoter DNA G-Quadruplexes. *J Am. Chem. Soc.* **130**, 15950–15956 (2008).
23. Nasiri, H. R. *et al.* Targeting a c-MYC G-quadruplex DNA with a fragment library. *Chem. Commun.* **50**, 1704–1707 (2014).
24. Chung, W. J., Heddi, B., Hamon, F., Teulade-Fichou, M.-P. & Phan, A. T. Solution Structure of a G-quadruplex Bound to the Bisquinolinium Compound Phen-DC3. *Angew. Chem. Int. Ed.* **53**, 999–1002 (2014).
25. Alzeer, J., Vummidi, B. R., Roth, P. J. & Luedtke, N. W. Guanidinium-modified phthalocyanines as high-affinity G-quadruplex fluorescent probes and transcriptional regulators. *Angew. Chem. Int. Ed.* **48**, 9362–9365 (2009).
26. Agarwal, T., Roy, S., Chakraborty, T. K. & Maiti, S. Selective targeting of G-quadruplex using furan-based cyclic homooligo peptides: effect on c-MYC expression. *Biochemistry* **49**, 8388–8397 (2010).
27. Chang, C.-C. *et al.* A fluorescent carbazole derivative: high sensitivity for quadruplex DNA. *Anal. Chem.* **75**, 6177–6183 (2003).
28. Jin, B. *et al.* Fluorescence light-up probe for parallel G-quadruplexes. *Anal. Chem.* **86**, 943–952 (2014).
29. Dash, J., Panda, D., Debnath, M. & Mandal, S. Anticancer activities of a heterocyclic molecule that selectively stabilizes G-quadruplex DNA. *Indian Patent*, Patent application no. 1136/2014 dated 5th November 2014.
30. Wu, Y., Guo, H., James, T. D. & Zhao, J. Enantioselective recognition of mandelic acid by a 3,6-dithiophen-2-yl-9H-carbazole-based chiral fluorescent bisboronic acid sensor. *J. Org. Chem.* **76**, 5685–5695 (2011).
31. Rostovtsev, V. V., Green, L. G., Fokin, V. V. & Sharpless, K. B. A stepwise Huisgen cycloaddition Process: copper(I)-catalyzed regioselective “ligation” of azides and terminal alkynes. *Angew. Chem. Int. Ed.* **41**, 2596–2599 (2002).
32. Darby, R. A. J. *et al.* High throughput measurement of duplex, triplex and quadruplex melting curves using molecular beacons and a LightCycler. *Nucleic Acids Res.* **30**, e39 (2002).
33. De Cian, A. *et al.* Fluorescence-based melting assays for studying quadruplex ligands. *Methods* **42**, 183–195 (2007).
34. Reichardt, C. Solvatochromic dyes as solvent polarity indicators. *Chem. Rev.* **94**, 2319–2358 (1994).
35. Das, R. N., Debnath, M., Gaurav, A. & Dash, J. Environment-sensitive probes containing a 2,6-diethynylpyridine motif for fluorescence turn-on detection and induction of nanoarchitectures of human telomeric quadruplex. *Chem. Eur. J.* **20**, 16688–16693 (2014).
36. White, E. W. *et al.* Structure-specific recognition of quadruplex DNA by organic cations: influence of shape, substituents and charge. *Biophys. Chem.* **126**, 140–153 (2007).
37. Balagurumoorthy, P. & Brahmachari, S. K. Structure and stability of human telomeric sequence. *J. Biol. Chem.* **269**, 21858–21869 (1994).
38. Jin, R., Gaffney, B. L., Wang, C., Jones, R. A. & Breslauer, K. J. Thermodynamics and structure of a DNA tetraplex: a spectroscopic and calorimetric study of the tetramolecular complexes of d(TG3T) and d(TG3T2G3T). *Proc. Natl. Acad. Sci. USA* **89**, 8832–8836 (1992).
39. Dash, J. *et al.* Synthesis of bis-indolecarboxamides as G-quadruplex stabilizing and inducing ligands. *Chem. Eur. J.* **18**, 554–564 (2012).
40. Mosmann, T. Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. *J. Immunol. Methods*, **65**, 55–63 (1983).
41. Steiner, P. *et al.* Identification of a Myc-dependent step during the formation of active G1 cyclin-cdk complexes. *EMBO J.* **14**, 4814–4836 (1995).
42. Dang, C. V. c-Myc target genes involved in cell growth, apoptosis, and metabolism. *MolCell Biol.* **19**, 1–11 (1999).
43. Boer, L. D. *et al.* Cyclin A/cdk2 coordinates centrosomal and nuclear mitotic events. *Oncogene* **27**, 4261–4268 (2008).
44. Boddupally, P. V. L. *et al.* Anticancer Activity and Cellular Repression of c-MYC by the G-Quadruplex-Stabilizing 11-Piperaziny lquindoline Is Not Dependent on Direct Targeting of the G-Quadruplex in the c-MYC Promoter. *J. Med. Chem.* **55**, 6076–6086 (2012).
45. Hwang, T. L. & Shaka, A. J. Water suppression that works. Excitation sculpting using arbitrary wave-forms and pulsed-field gradients. *J. Magn. Res. Series A.* **112**, 275–279 (1995).
46. Sklenar, V. & Bax, A. Spin-echo water suppression for the generation of pure-phase two-dimensional NMR spectra. *J. Magn. Res.* **74**, 469–479 (1987).
47. Wu, C.-H. *et al.* Cellular senescence is an important mechanism of tumor regression upon c-Myc inactivation. *Proc. Natl. Acad. Sci. USA.* **104**, 3028–3033 (2007).
48. Veliça, P. *et al.* Lack of functional and expression homology between human and mouse aldo-ketoreductase 1C enzymes: implications for modeling human cancers. *Mol Cancer.* **8**, 121 (2009).
49. Leite, F. *et al.* Inflammatory cytokines enhance the interaction of mannheimia haemolyticaleukotoxin with bovine peripheral blood neutrophils *in vitro*. *Infect. Immun.* **70**, 4336–4343 (2002).
50. Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. Protein measurement with the Folin phenol reagent. *J. Biol. Chem.* **193**, 265–275 (1951).

Acknowledgments

This work was supported by the Department of Science and Technology (DST), Council of Scientific & Industrial Research (CSIR), India and Department of Biotechnology (DBT) India for funding. DP thanks CSIR, India and MD thanks DST, India for research fellowships. HS and IB thank the state of Hesse (BMRZ, LOEWE project: Synchembio) for financial support.

Author Contributions

J.D. designed the experiments, D.P. synthesized the carbazole ligands, D.P. and M.D. carried out FRET melting, UV/Vis, fluorescence and CD experiments. M.D. and S.M. performed biological experiments. I.B. and H.S. are responsible for the N.M.R. analysis. J.D. and M.D. wrote the paper. H.S. read the paper and gave overall comments on the paper.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Panda, D. *et al.* A Nucleus-Imaging Probe That Selectively Stabilizes a Minor Conformation of c-MYC G-quadruplex and Down-regulates c-MYC Transcription in Human Cancer Cells. *Sci. Rep.* **5**, 13183; doi: 10.1038/srep13183 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

6.3 Research article: A fluorescent guanosine dinucleoside as a selective switch-on sensor for *c-myc* G-quadruplex DNA with potent anticancer activities

Kumar Y. P., Bhowmik S., Das R. N., Bessi I., Paladhi S., Ghosh R., Schwalbe H., Dash J. Chemistry, 2013, 26; 19(35): 11502-6.

This article explores the potential applications of a novel fluorescent diguanosine dansyl derivative (DDG) as G-quadruplex ligand.

DDG revealed to be highly selective for the *c-MYC* G-quadruplex. Binding to *c-MYC* resulted in a remarkable enhancement of fluorescence emission compared to other promoter G-quadruplexes and duplex DNA. The ligand binding mode to *c-MYC* was investigated by NMR and the chemical shift perturbations produced by the ligand were mapped on the *c-MYC* structure. Fluorescence microscopy revealed that DDG is cell-permeable and binds preferentially to the nucleus. Furthermore, the ligand is able to arrest the cell cycle at the G2/M phase and to inhibit cell growth and induce apoptosis in A375 cancer cells with selectivity over normal cells.

The design and the synthesis of the ligand DDG, the fluorescence studies as well as all the biological essays were carried out in the group of Prof. Dash (IISER Kolkata, India). The author of this thesis performed the NMR experiments on the interaction of the *c-MYC* quadruplex with DDG, analysed the NMR data and wrote the corresponding paragraph in the article.

A Fluorescent Guanosine Dinucleoside as a Selective Switch-On Sensor for *c-myc* G-Quadruplex DNA with Potent Anticancer Activities

Y. Pavan Kumar,^[a] Sudipta Bhowmik,^[a] Rabindra Nath Das,^[b] Irene Bessi,^[c]
Sushovan Paladhi,^[a] Rita Ghosh,^[d] Harald Schwalbe,^[c] and Jyotirmayee Dash*^[a, b]

Dedicated to Professor Krishna N. Ganesh on the occasion of his 60th birthday

Development of fluorescent chemical probes that can recognize biomacromolecules with high specificity and interfere with cellular processes is an emerging trend in chemical biology.^[1,2] Within this context, G-quadruplex DNA has received considerable attention as a prospective target for designing anticancer drugs.^[3–4] G-Quadruplex structures are widespread in the genome, found at the end of telomeres, in the promoter regions proto-oncogenes, and in the untranslated regions of mRNAs. These structures are believed to play key roles in the human genome, such as telomere maintenance and gene regulation.^[4] While several classes of molecules have been reported for stabilization of G-quadruplex DNA, only a few fluorescent chemosensors for the selective detection of G-quadruplex motifs have been developed.^[5,6] In addition, there is current interest in developing fluorescent probes with multiple signals that can find applications in analytical and computational devices.^[1–2,7] Recently we have shown that it is possible to create biomolecular logic-gate systems based on the interaction of fluorescent molecular probes with the G-quadruplex DNA.^[5f]

Monchaud and co-workers have reported the design and synthesis of two molecules containing guanine bases and their possible self-assembly to form artificial G-quartets as a nature-inspired strategy to interact with the G-quadruplex.^[8] These ligands show selectivity for quadruplexes over duplex

DNA. However, the G-quartets (self-assembled Hoogsteen-type hydrogen-bonded macrocycles of four guanine bases) are common to all quadruplexes, making discrimination between the quadruplexes challenging.

Previously we have reported synthesis of G-quadruplex binding ligand^[9] using “click-chemistry”^[9c] and fabrication of complex nanoarchitectures using supramolecular self-assembly of guanosine derivatives.^[10] Among the five natural nucleobases, guanine and its derivatives have received considerable interest in supramolecular chemistry and nanotechnology.^[11] Taking inspiration from the natural self-organization of nucleosides, we envisioned designing a flexible ligand by linking a biocompatible fluorescent tag between two guanosine units using Cu^I-catalyzed 1,3-dipolar azide-alkyne cycloaddition.^[12,13] We hypothesized that the guanosine units can interact with the G-quartet of the quadruplex by means of hydrogen bonding and the flexible linker can tether in the groove region of quadruplex sequences. Since quadruplex sequences vary in the groove and loop regions, the ligand may show selectivity for a particular quadruplex sequence, for which the groove region would be maximally occupied. Further the ligand based on a guanosine nucleoside can be incorporated into DNA and the attached fluorophore would enable visualization of the nucleus in living cells.^[14]

Based on our design principle, we have synthesized a guanosine azide **1** from the natural nucleoside guanosine in three steps and a biocompatible fluorescent dansyl probe **2** (Scheme 1 and Scheme S1 in the Supporting Information). The C₂-symmetric dinucleoside **3** (DDG) was prepared in high yields by employing a double click reaction of **1** with the dansyl dialkyne **2** in the presence of Na ascorbate and CuSO₄·5H₂O in *t*BuOH/H₂O (1:1).

The ability of DDG to discriminate nucleic acid sequences was investigated using fluorescence spectroscopy (Figure 1). The emission spectra of the guanosine–dansyl conjugate DDG was characterized by a twofold intense peak at 430 nm and a minor peak at 553 nm, when excited at 350 nm (quantum yield, $\Phi=0.25$). Guanosine azide **1** is essentially non-fluorescent; however, the emission bands of DDG at 430 and 553 nm were assigned to the guanosine and the dansyl group, respectively, owing to the excitation energy transfer from the guanosine to the dansyl chromo-

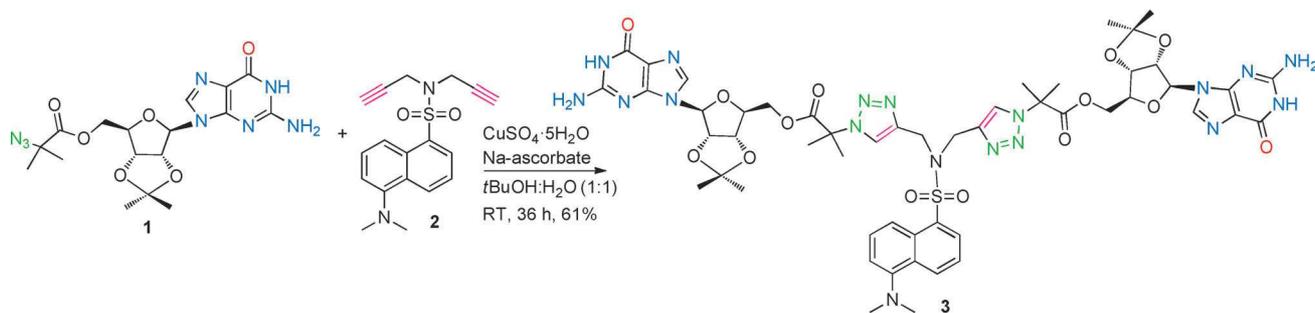
[a] Y. P. Kumar, Dr. S. Bhowmik, S. Paladhi, Prof. Dr. J. Dash
Department of Organic Chemistry
Indian Association for the Cultivation of Science Jadavpur
Kolkata 700032 (India)
E-mail: ocjd@iacs.res.in

[b] R. N. Das, Prof. Dr. J. Dash
Department of Chemical Sciences
Indian Institute of Science Education and Research Kolkata
Mohanpur, West Bengal 741252 (India)

[c] I. Bessi, Prof. Dr. H. Schwalbe
Institute of Organic Chemistry and Chemical Biology
Goethe University Frankfurt and
Center for Biomolecular Magnetic Resonance
Max-von-Laue Strasse 7, 60438 Frankfurt am Main (Germany)

[d] R. Ghosh
Department of Biochemistry and Biophysics
University of Kalyani, Kalyani-741235, West Bengal (India)

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/chem.201302107>.



Scheme 1. Modular synthesis of a fluorescent guanosine dinucleoside **3** (DDG).

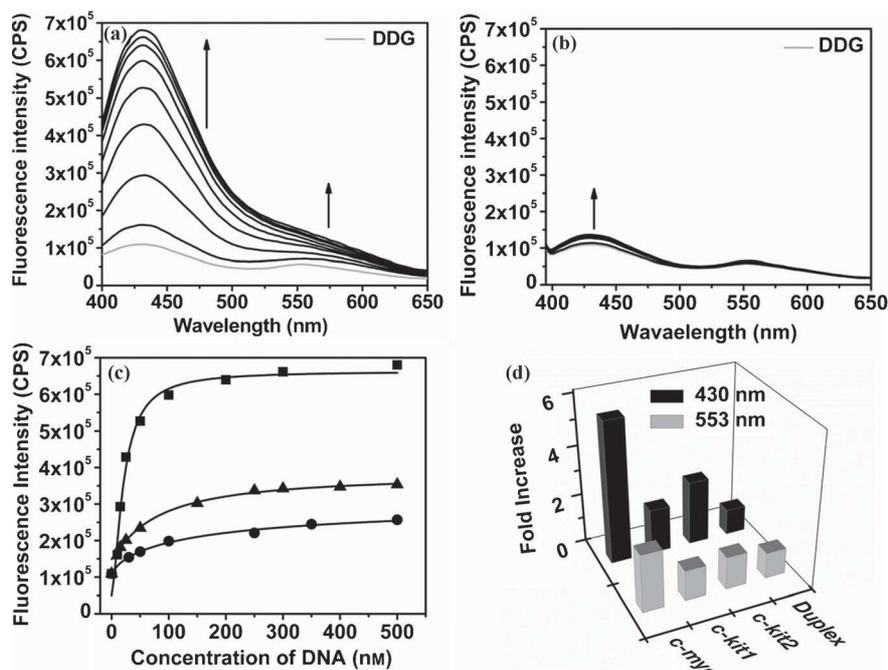


Figure 1. Fluorimetric titration spectra of DDG (500 nm) with a) *c-myc* quadruplex DNA (0–500 nm) and b) with duplex DNA (0–1500 nm). c) Fluorescence intensity profiles of DDG (500 nm) with quadruplex sequences (0–500 nm) at 430 nm (■ = *c-myc*, ● = *c-kit1*, ▲ = *c-kit2*). d) Fold increase of fluorescence intensity for DDG (500 nm) upon titration with DNA sequences (100 nm).

phore.^[15,16] This indicates that DDG, which contains biologically active non-natural and natural components, such as dansyl sulphonamide, triazole and guanosine, exhibits distinct optical properties and may lead to enhance biological activities.

When the fluorimetric titration of DDG (500 nm) was carried out at pH 7.4 with increasing concentration of quadruplex sequences (0–500 nm), the fluorescence intensity of both dansyl and guanosine band was enhanced in a dose-dependent manner from which the dissociation binding constants (K_d) were calculated (Figure 1 and Figures S2 and S3 and Table S1 in the Supporting Information).^[17] In the emission spectra, the addition of *c-myc* resulted higher selectivity, with a 6-fold enhancement of guanosine band and 2.5-fold enhancement for the dansyl band. It shows high selectivity for *c-myc* over *c-kit2* and *c-kit1* quadruplexes and

duplex DNA (Figure 1d). A negligible fluorescence enhancement of DDG was observed for duplex DNA at both the guanosine and dansyl bands. By comparing the change in ratio of the fold increase between guanosine and dansyl emission of DDG upon titration with DNA sequences, it was observed that in all cases the guanosine band was increased from one- to fourfold compared to the dansyl band. The ratiometric variation of DDG was dependent upon the nucleic acid sequences and recognition occurred at very low concentration of DNA. Interestingly, even in the presence of 100 nm *c-myc*, the emission intensity of DDG was increased fivefold at 430 nm (Figure 1d). It is important to point out that such nanomolar sensitivity is highly impressive for the development of biomarker for the quadruplex DNA. To the best

of our knowledge, none of the previously reported probes can detect nanomolar DNA with high selectivity.^[5] DDG showed both excellent affinity for *c-myc* quadruplex ($K_d = 20.8$ nm) as well as substantial specificity (44.5-fold) over duplex DNA ($K_d = 925.7$ nm) at 430 nm. DDG also exhibited 3.8-fold selectivity for *c-myc* over *c-kit2* ($K_d = 79.4$ nm) and 7.6-fold selectivity for *c-myc* over *c-kit1* ($K_d = 158.8$ nm).

Based on the high specificity of the guanosine–dansyl conjugate for *c-myc* quadruplex DNA, we performed excited-state lifetime measurements^[18] of DDG in the absence and presence of *c-myc* quadruplex DNA (Figure 2). The results reveal that DDG has two forms in the excited state, a major component (95.5%) with a lifetime of 454 ps and a minor form with a higher lifetime (1.7 ns). Upon addition of 1.0 equivalents of *c-myc* quadruplex, an enhancement in the excited state lifetime of both the species was observed.

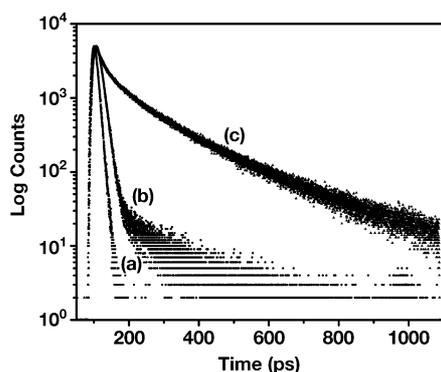


Figure 2. Fluorescence decay profiles ($\lambda_{\text{ex}} = 340$ nm) of a) the excitation lamp profile, b) DDG (500 nm), and c) DDG and *c-myc* (500 nm).

However, the decay profiles show that the minor form becomes the major component (75%) with a higher lifetime of 6.4 ns (Table S2, in the Supporting Information). These results indicate that DDG binds to *c-myc* G-quadruplex DNA and forms a stable complex with higher fluorescence as well as higher lifetime.

The interaction of DDG with G-quadruplex DNA was further evaluated by Förster resonance energy transfer (FRET) melting analysis^[19] using dual-labeled DNA sequences (Figure S4 in the Supporting Information). FRET melting analysis has been used to determine the G-quadruplex stabilization and selectivity of a ligand towards a particular G-quadruplex target. We have evaluated the melting of four DNA sequences with G-rich DNA sequences found in the promoter of *c-myc*, two G-quadruplexes found in the promoter of *c-kit* (*c-kit1* and *c-kit2*), and a self-complementary duplex DNA sequence. FRET melting studies showed good agreement with the fluorescence binding titrations (Table S3 in the Supporting Information). The stabilization potential for *c-myc* saturates at 1 μM concentration of DDG ($\Delta T_m = 12.6$ K at 1 μM , that is, a T_m of 93°C). DDG also showed good stabilization for *c-kit* quadruplex sequences with a ΔT_m of 17.5 K for *c-kit1* and ΔT_m of 7.4 K for *c-kit2* at a DDG concentration of 1 μM , whereas no detectable duplex stabilization ($\Delta T_m = 0.5^\circ\text{C}$) was observed for the DDG at this concentration.

To shed light on the binding mode, the titration of DDG with *c-myc* DNA was monitored by 1D ^1H NMR spectra (Figure 3). In the region of imino proton signals, NMR resonances from G13, G14/G4 (overlapped signals) residues revealed significant line broadening. Also in the aromatic region, changes in chemical shifts could not be quantified for all residues due to signal overlap. Among the monitored aromatic protons, residues G2, A12, G13, and A22 showed the most pronounced chemical-shift perturbations. Further, in the ^1H , ^1H -NOESY spectrum (Figure 3c), the cross-peak of imino signal of G6 to G5 and of G6* to G5* in the second minor conformation of *c-myc*^[20] is shifted, suggesting that G6 is interacting with the ligand. Furthermore, the cross-peaks of imino signal of G13 to G8, G9 and G14 are much weaker in the presence of the ligand than in the free

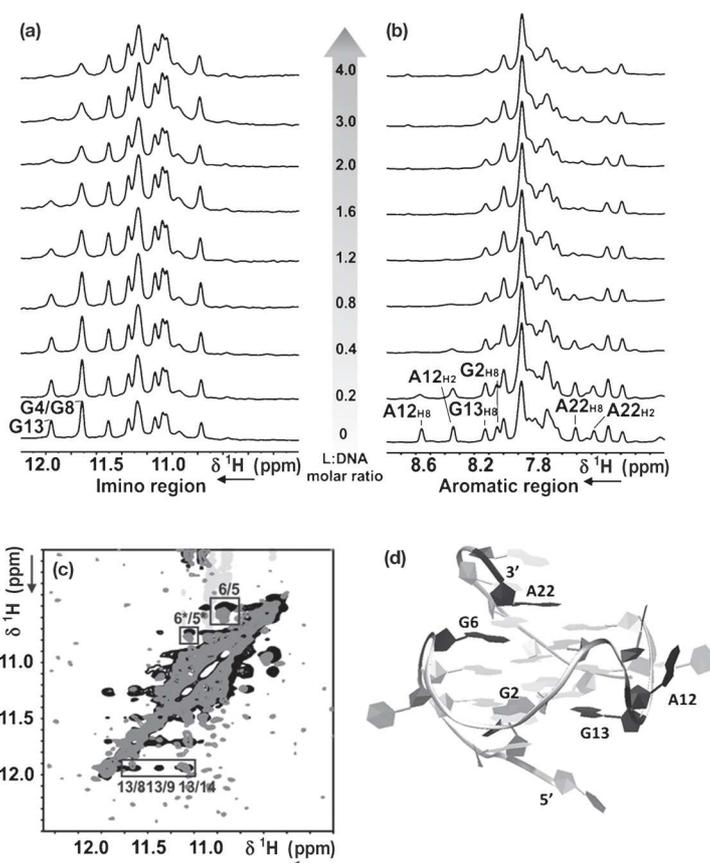


Figure 3. NMR spectra (600 MHz) of a) the imino region and b) the aromatic region of 200 μM *c-myc* in the presence of DDG at different DDG:*c-myc* DNA molar ratio. c) Overlay of imino region of ^1H , ^1H -NOESY of the complex (grey: *c-myc*:DDG=1:2, 300 μM *c-myc*, 800 MHz) with the DNA alone (black: 1 mM *c-myc*, 950 MHz); Titration was performed at 298 K, in buffer containing 25 mM tris-HCl at pH 7.4 and 100 mM KCl, $\text{H}_2\text{O}/[\text{D}_6]\text{DMSO}$ 80:20. d) Map of chemical shift perturbations detected by NMR spectroscopy on the *c-myc* model (PDB: 1XAV); adenosine and guanine perturbed residues in black.

from, confirming that G13 must be involved in the interaction with DDG, in line with the NMR data from titration experiments. Our initial NMR analysis therefore shows that DDG interacts with both the external G-quartets and with the groove defined by G8-G9-G10 and G13-G14-G15 (Figure 3d).

Human malignant melanoma A375 cells were used as a model to investigate the cell membrane permeability of DDG (Figure 4). The cells were incubated with DDG (2.5 μM) for 30 min at 37°C and examined by fluorescence microscopy. It was found that DDG, which contains the guanosine nucleoside unit and hydrophobic heteroaromatic core, is cell-membrane permeable and displays blue fluorescence (Figure 4b). The images indicated that DDG binds to the nucleus preferentially, but it also stains other parts of the cell. This result was further confirmed by co-staining DDG with the nucleus marker propidium iodide (PI; Figure 4c).

The antiproliferative activities of the dansyl-guanosine conjugate DDG was then determined on the A375 cell line.

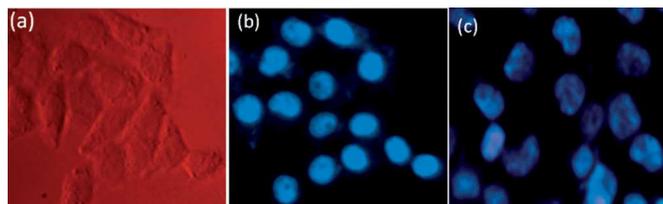


Figure 4. Fluorescence microscope images of A375 cells. a) Bright-field transmission image of cells after incubation with DDG (2.5 μM) at 37°C for 30 min. b) Fluorescence image of a). c) Fluorescence image of DDG stained (same as a) fixed with 4% paraformaldehyde and stained with PI (1.0 μM) for 30 min.

The cytotoxic effect of DDG on A375 cells was determined with varying concentrations (0, 0.5, 1, 10, 25, 50 μM) for 24 h by MTT assay (MTT = 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide).^[21] The IC_{50} value of DDG was found to be 5 μM (Figure S5 in the Supporting Information). Despite this potency, DDG exhibited lower cytotoxicity towards human normal keratinocyte (HaCaT) cells with an IC_{50} value at 69 μM . These results suggest that DDG possess significant selectivity for cancer cells over normal cells. Next we have investigated the ability of DDG to arrest the cell cycle and induce apoptosis using A375 cells by fluorescence-activated cell sorting (FACS).^[22–24] The cell cycle analysis revealed that DDG can disturb the cell cycle of the A375 cell line after exposure for only 12 h, which is shown with the decreased G0–G1 phase and an increased G2/M phase in a dose-dependent manner (Figure S6 in the Supporting Information). The treatment of DDG caused cell cycle arrest in G2/M phase,^[25] possibly due to repair of damaged DNA phase.

To determine whether the observed cell death induced by DDG was due to apoptosis, a biochemical marker of apoptosis, for example, mitochondrial membrane depolarization (Ψ_m), was monitored by flow cytometry after JC-1 staining.^[22–24] As shown in Figure 5, untreated cells have intact mitochondria and normal $\Delta\Psi_m$; however, the loss of $\Delta\Psi_m$ exhibits a dose-dependent increase after the cells were treated with DDG, as evidenced by the shift of fluorescence of the JC-1 dye from red to green. DDG at the concentrations of 1, 2.5 and 5 μM for 24 h increased $\Delta\Psi_m$ of cells from 8.36 to 39.06, 46.27, and 50.27%, respectively. These experiments indicated that cancer cells treated by DDG lose $\Delta\Psi_m$ and induce apoptosis in A375 cells through mitochondria-mediated pathways. The viability of A375 cells treated with DDG was further investigated by PI-flow cytometric analysis.^[24] The in vitro apoptosis analysis indicated that exposure of A375 cells to different concentrations of DDG for 24 h resulted in a dose-dependent increase in the proportion of apoptotic cells (16.06 to 36.11%, Figure S7 in the Supporting Information).

We have shown here the first example of a synthetic guanosine-based novel fluorescent biosensor that shows rapid, selective, and sensitive turn-on responses to nucleic acid sequences. The fluorescence enhancement of the dansyl guanosine derivative in the presence of *c-myc* DNA occurs in

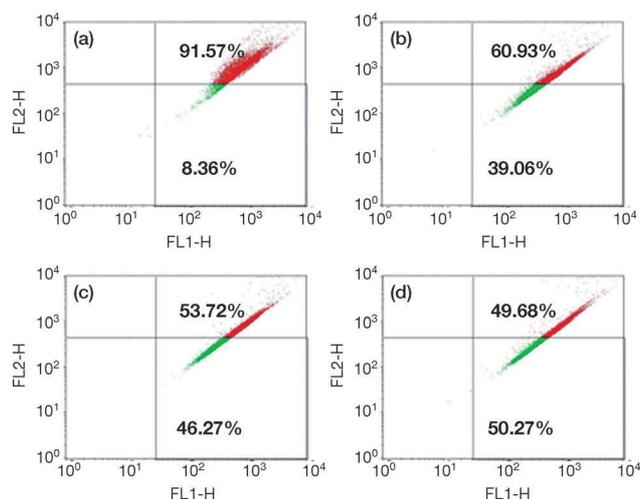


Figure 5. Flow cytometric analysis of the change in mitochondrial potential. A375 cancer cells were incubated for 24 h in the presence of (0–5 μM) of DDG using JC1-stained cells. Histograms a) 0 μM , b) 1.0 μM , c) 2.5 μM , and d) 5.0 μM indicate the percent of normal (M1) and apoptotic cells (M2) upon treatment with DDG.

real time, which could potentially be developed as a biomarker for the quadruplex in the nanomolar concentration range. Dinucleoside DDG induces cell cycle arrest at G2/M phase, exhibits specific cytotoxicity to the human melanoma A375 cells over normal human keratinocyte cells, and promotes cell death by apoptosis.

Acknowledgements

This work was supported by the Department of Science and Technology (DST), Department of Biotechnology (DBT) and Council of Science and Industrial Research (CSIR), India. R.N.D. and S.P. thank CSIR India for research fellowships. The authors thank Prof. Dr. Shankar Balasubramanian, University of Cambridge, and Prof. Dr. Kankan Bhattacharyya, IACS Kolkata, for helpful comments on the manuscript.

Keywords: cell cycles • click chemistry • fluorescence • nucleosides • G-quadruplex DNA

- [1] a) A. W. Czarnik, *Fluorescent Chemosensors for Ion and Molecule Recognition*, American Chemical Society, Washington, DC, **1992**; b) A. P. de Silva, in *Molecular and Supramolecular Chemistry. Vol. 7: Supramolecular Devices* (Eds.: J. W. Steed, P. A. Gale, P. Anzenbacher), Wiley-Blackwell, Oxford, **2012**, p. 2497.
- [2] a) A. P. de Silva, S. Uchiyama, *Nat. Nanotechnol.* **2007**, *2*, 399–410; b) M. S. T. Gonçalves, *Chem. Rev.* **2009**, *109*, 190–212; c) R. W. Sinkeldam, N. J. Greco, Y. Tor, *Chem. Rev.* **2010**, *110*, 2579–2619; d) R. N. Dsouza, U. Pischel, W. M. Nau, *Chem. Rev.* **2011**, *111*, 7941–7980; e) X. Chen, X. Tian, I. Shin, J. Yoon, *Chem. Soc. Rev.* **2011**, *40*, 4783–4804; f) X. Chen, T. Pradhan, F. Wang, J. S. Kim, J. Yoon, *Chem. Rev.* **2012**, *112*, 1910–1956; g) M. Vendrell, D. Zhai, J. Cheng, Y. T. Chang, *Chem. Rev.* **2012**, *112*, 4391–4420.
- [3] S. Neidle, S. Balasubramanian, *Quadruplex Nucleic Acids*, Royal Society of Chemistry, Cambridge (UK), **2006**.
- [4] a) J. T. Davis, *Angew. Chem.* **2004**, *116*, 684–716; *Angew. Chem. Int. Ed.* **2004**, *43*, 668–698; b) D. Monchaud, M. P. Teulade-Fichou, *Org.*

- Biomol. Chem.* **2008**, *6*, 627–636; c) Y. Qin, L. H. Hurley, *Biochimie* **2008**, *90*, 1149–1171; d) S. Balasubramanian, S. Neidle, *Curr. Opin. Chem. Biol.* **2009**, *13*, 345–353; e) S. Balasubramanian, L. H. Hurley, S. Neidle, *Nat. Rev. Drug Discovery* **2011**, *10*, 261–275; f) G. W. Collie, G. N. Parkinson, *Chem. Soc. Rev.* **2011**, *40*, 5867–5892; g) M. L. Bochman, K. Paeschke, V. A. Zakian, *Nat. Rev. Genet.* **2012**, *13*, 770–780.
- [5] For a review, see: a) E. Lary, A. Granzhan, F. Hamon, D. Verga, M.-P. Teulade-Fichou, *Top. Curr. Chem.* **2012**, *330*, 111–178. For selected examples, see: b) F. Koepfel, J. F. Riou, A. Laoui, P. Mailliet, P. B. Arimondo, D. Labit, O. Petitgenet, C. Helene, J. L. Mergny, *Nucleic Acids Res.* **2001**, *29*, 1087–1096; c) P. Yang, A. De Cian, M. P. Teulade-Fichou, J. L. Mergny, D. Monchaud, *Angew. Chem.* **2009**, *121*, 2222–2225; *Angew. Chem. Int. Ed.* **2009**, *48*, 2188–2191; d) M. Nikan, M. Di Antonio, K. Abecassis, K. McLuckie, S. Balasubramanian, *Angew. Chem. Int. Ed.* **2013**, *52*, 1428–1431; e) J. Mohanty, N. Barooah, V. Dhamodharan, S. Hari Krishna, P. I. Pradeepkumar, A. C. Bhasikuttan, *J. Am. Chem. Soc.* **2013**, *135*, 367–376; f) S. Bhowmik, R. N. Das, B. Parasar, J. Dash, *Chem. Commun.* **2013**, *49*, 1817–1819; and references there in.
- [6] For selected metal complexes, see: a) K. Suntharalingam, Łęczkowska, M. A. Furrer, Y. Wu, M. K. Kuimova, B. Therrien, A. J. P. White, R. Vilar, *Chem. Eur. J.* **2012**, *18*, 16277–16282; b) J. Alzeer, B. R. Vummidi, P. J. C. Roth, N. W. Luedtke, *Angew. Chem.* **2009**, *121*, 9526–9529; *Angew. Chem. Int. Ed.* **2009**, *48*, 9362–9365; and references there in.
- [7] H.-Z. He, D. S.-H. Chan, C.-H. Leung, D.-L. Ma, *Nucleic Acids Res.* **2013**, *41*, 4345–4359.
- [8] a) L. Stefan, A. Guédin, S. Amrane, N. Smith, F. Denat, J. L. Mergny, D. Monchaud, *Chem. Commun.* **2011**, *47*, 4992–4994; b) H. J. Xu, L. Stefan, R. Haudecoeur, S. Vuong, P. Richard, F. Denat, J. M. Barbe, C. P. Gros, D. Monchaud, *Org. Biomol. Chem.* **2012**, *10*, 5212–5218; c) R. Haudecoeur, L. Stefan, F. Denat, D. Monchaud, *J. Am. Chem. Soc.* **2013**, *135*, 550–553.
- [9] a) J. Dash, P. S. Shirude, S. Balasubramanian, *Chem. Commun.* **2008**, 3055–3057; b) J. Dash, P. S. Shirude, S. T. D. Hsu, S. Balasubramanian, *J. Am. Chem. Soc.* **2008**, *130*, 15950–15956; c) J. Dash, Z. A. E. Waller, G. D. Pantoş, S. Balasubramanian, *Chem. Eur. J.* **2011**, *17*, 4571–4581; d) J. Dash, R. N. Das, N. Hegde, G. D. Pantoş, P. S. Shirude, S. Balasubramanian, *Chem. Eur. J.* **2012**, *18*, 554–564.
- [10] a) J. Dash, A. J. Patil, R. N. Das, F. Dowdal, S. Mann, *Soft Matter* **2011**, *7*, 8120–8126; b) R. N. Das, Y. P. Kumar, S. Pagoti, A. J. Patil, J. Dash, *Chem. Eur. J.* **2012**, *18*, 6008–6014.
- [11] a) J. T. Davis, G. P. Spada, *Chem. Soc. Rev.* **2007**, *36*, 296–313; b) S. Lena, S. Masiero, S. Pieraccini, G. P. Spada, *Chem. Eur. J.* **2009**, *15*, 7792–7806.
- [12] V. V. Rostovtsev, L. G. Green, V. V. Fokin, K. B. Sharpless, *Angew. Chem.* **2002**, *114*, 2708–2711; *Angew. Chem. Int. Ed.* **2002**, *41*, 2596–2599.
- [13] a) F. Amblard, J. H. Cho, R. F. Schinazi, *Chem. Rev.* **2009**, *109*, 4207–4220; b) H. C. Kolb, K. B. Sharpless, *Drug Discovery Today* **2003**, *8*, 1128–1137.
- [14] Y. N. Teo, E. T. Kool, *Chem. Rev.* **2012**, *112*, 4221–4245.
- [15] J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*, 2nd ed., Kluwer Academic/Plenum, New York, **1999**.
- [16] For PET based RNA sensors, see; B. A. Sparano, K. Koide, *J. Am. Chem. Soc.* **2007**, *129*, 4785–4794.
- [17] H. M. Kim, M. S. Seo, M. J. An, J. H. Hong, Y. S. Tian, J. H. Choi, O. Kwon, K. J. Lee, B. R. Cho, *Angew. Chem.* **2008**, *120*, 5245–5248; *Angew. Chem. Int. Ed.* **2008**, *47*, 5167–5170.
- [18] M. Y. Berezin, S. Achilefu, *Chem. Rev.* **2010**, *110*, 2641–2684.
- [19] a) R. A. J. Darby, M. Sollogoub, C. McKeen, L. Brown, A. Risitano, N. Brown, C. Barton, T. Brown, K. R. Fox, *Nucleic Acids Res.* **2002**, *30*, 39e; b) A. De Cian, L. Guittat, M. Kaiser, B. Sacca, S. Amrane, A. Bourdoncle, P. Alberti, M.-P. Teulade-Fichou, L. Lacroix, J. L. Mergny, *Methods* **2007**, *42*, 183–195.
- [20] a) A. Ambrus, D. Chen, J. Dai, R. A. Jones, D. Yang, *Biochemistry* **2005**, *44*, 2048–2058; b) J. Dai, M. Carver, L. H. Hurley, D. Yang, *J. Am. Chem. Soc.* **2011**, *133*, 17673–17680.
- [21] T. Mosmann, *J. Immunol. Methods* **1983**, *65*, 55–63.
- [22] C. Riccardi, I. Nicoletti, *Nat. Protoc.* **2006**, *1*, 1458–1461.
- [23] T. Liu, B. Hannafon, L. Gill, W. Kelly, D. Benbrook, *Mol. Cancer Ther.* **2007**, *6*, 1814–1822.
- [24] M. van Gurp, N. Festjens, G. van Loo, X. Saelens, P. Vandenabeele, *Biochem. Biophys. Res. Commun.* **2003**, *304*, 487–497.
- [25] For a G-quadruplex stabilizer that induces M-phase cell cycle arrest; see: Y.-C. Tsai, H. Qi, C.-P. Lin, R.-K. Lin, J. E. Kerrigan, S. G. Rzuczek, E. J. Lavoie, J. E. Rice, D. S. Pilch, Y. L. Lyu, L. F. Liu, *J. Biol. Chem.* **2009**, *284*, 22535–22543.

Received: June 3, 2013

Published online: July 29, 2013

7 Insight into G-quadruplex folding kinetics by real-time NMR

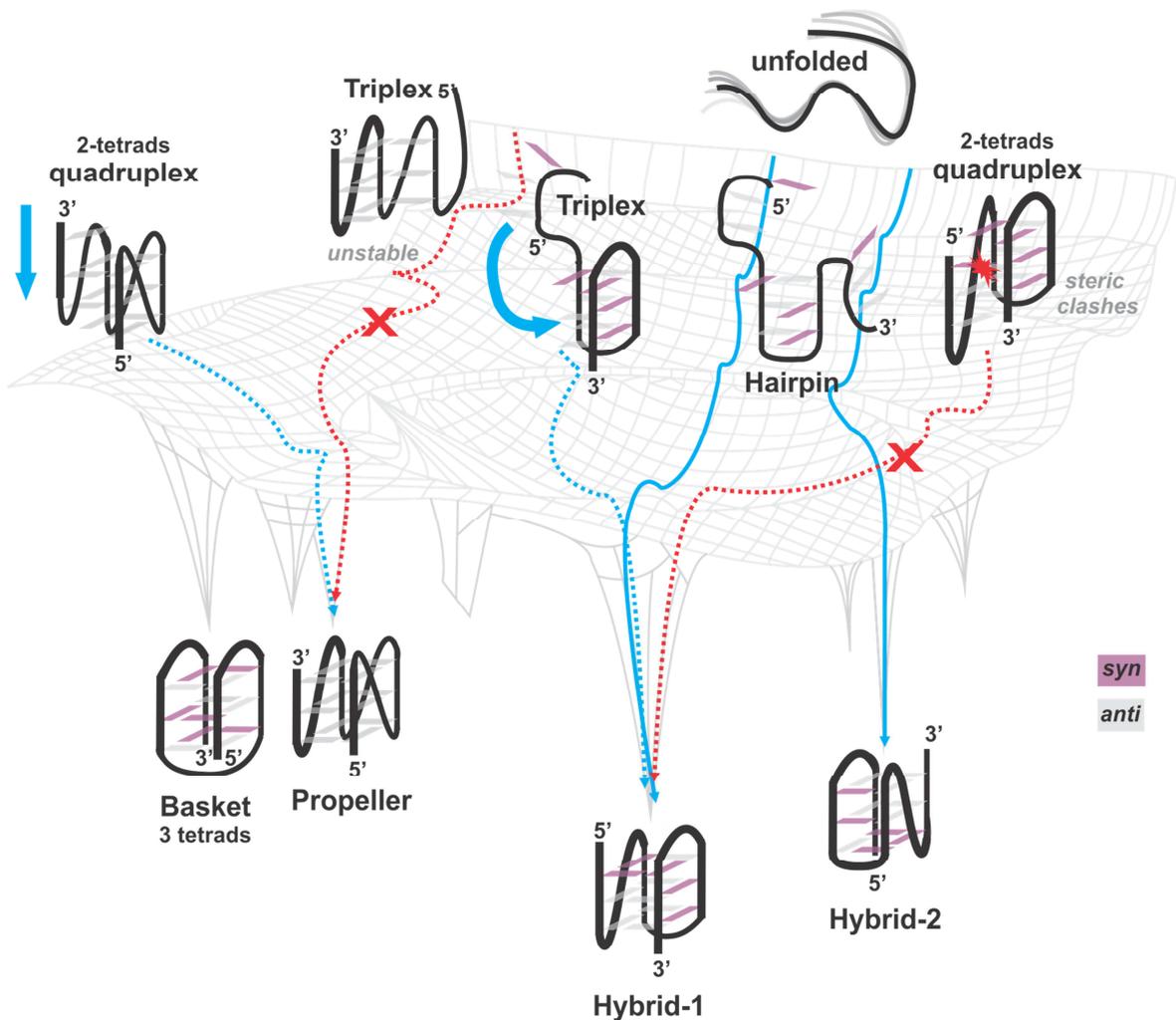


Figure 44 Sketch of a possible energy landscape for the human telomeric G-quadruplex, inspired from reference [398]. The structures that have been experimentally characterized and described in Chapter 2.4 occupy the deepest basins, while the proposed intermediate states (triplex, two-tetrad quadruplex and hairpin) may be local minima with overall higher energy and more or less shallow surroundings. *Anti* and *syn* guanine residues are represented in gray and magenta, respectively. The possible folding pathways are indicated with cyan arrows, while the forbidden ones are depicted in red. The proposed folding pathway of Tel24 (see Chapter 7.1 for discussion) is indicated with full line.

The energy landscape of the human telomeric G-quadruplex is most likely very rugged and features many competing basins of attraction. The relative levels of the energy minima can be adjusted by fine tuning the experimental conditions (cation nature and concentration, temperature, flanking nucleotides).

Figure 44 depicts as an example an arbitrary scenario for the human telomeric G-quadruplex energy landscape.

Investigation of the folding kinetics of G-quadruplex structures is essential to understand the interaction between telomere-binding proteins (see also Chapter 3.1) and the 3'-overhang of telomeres. Using the knowledge on the folding pathways, G-quadruplex ligands, able to bind selectively to the conformations that are actually present on a relevant biological time scale, can be developed.

While the diverse telomeric G-quadruplexes have been rather well characterized from a structural point of view (see Chapter 2.4), much less is known about their folding kinetics.

In the scientific community there is no general agreement on the folding pathway of monomolecular G-quadruplexes, regarding for example the timescale of the process as well as the number and the structural nature of the species involved in the folding. The discrepancies are usually remarkable when the kinetic data extracted from different experimental techniques are compared. For example, the kinetic of folding of the telomeric sequence Tel22 d[AGGG(TTAGGG)₃] has extensively been investigated in Chaires' group using the stopped-flow rapid mixing technique. In early studies, the progress of the folding was monitored by UV absorbance and it was suggested that the K⁺-induced folding of human telomeric quadruplex was completed within 10 ms after mixing.^[399] However, the same process monitored by 2-aminopurine (2AP) fluorescence exhibited relaxation constants in the range 40-900 s.^[78] In 2014, still in Chaires' group, analysis of the folding monitored *via* CD provided relaxation times in the order of 10⁴ s,^[400] suggesting that the folding of telomeric quadruplex can be very slow, in line with what we observed by real-time NMR (see Chapter 7.1). These striking differences may be ascribed to the insensitivity of UV and CD spectroscopy to monitor the structural features of the populations of the different states during the folding. For example, the observed CD signal is intrinsically ambiguous due to the fact that different conformations can contribute to the ellipticity at the wavelength selected for the kinetic data analysis. Structures with different loops arrangement, but same relative orientation of the G-stretches (*e.g.*, hybrid-1 and hybrid-2) produce similar CD spectra.^[401] On the other side, similar G-quadruplex topologies can be featured by significantly different CD signature.^[82] In fact, atomic-resolution techniques such as NMR are necessary to unravel the variety of structures that can be populated along the G-quadruplex folding pathway and to provide more correct and complete kinetic data. Not only the method chosen for the kinetic investigation, but also the experimental conditions play a major role in the differences reported in the various kinetic studies. In particular, the molar ratio of K⁺/DNA, the presence of flanking nucleotides and the annealing method can largely affect the ruggedness of the energy surface and the distribution of the energy minima.

Irrespective of these discrepancies, all the studies reported until now suggest that the folding of telomeric G-quadruplex is multiphasic and involves intermediates.

The folding pathway of the human telomeric G-quadruplex may be described by the kinetic partitioning mechanism as proposed for protein and, more recently, for RNA.^[402] While a fraction of the molecules is able to reach directly the native conformation, the rest can be kinetically trapped in other non-native basins of the energy landscape. The recent finding that G-quadruplex folding is very slow is in fact supporting the idea of a very rugged energy surface, with deep competing

energy minima corresponding to metastable conformations that can be populated during the folding. Interestingly, Bevilacqua and co-workers proposed that an increase in the length of the G-stretches can be used to deliberately populate more intermediate states and reduce the cooperativity of the folding.^[403]

A variety of intermediates has been proposed over the years on the basis of different experimental techniques: two-tetrads quadruplexes, triplexes and hairpins. None of these structures has been supported by clear experimental evidence. As already mentioned, CD signatures and UV absorbance are not sensitive to the specific quadruplex conformation and any intermediate structure proposed on the basis of CD and UV data is purely speculative. Nevertheless, there are recent reports where spectra of rapidly formed intermediate species have been calculated by singular value decomposition (SVD) of CD stopped flow data and assigned to specific structures.^[400]

Real-time NMR is a valuable tool to identify the species involved in the folding pathway at atomic resolution. Even though the folding pathway may be simplified because of the time-resolution, we have used real-time NMR to investigate the folding of telomeric G-quadruplex and the results of these studies are presented in the next Chapter.

It is clear however that, due to the time resolution, real-time NMR alone cannot provide a complete description of the folding and other atomic-resolution techniques have to be used in addition to obtain complementary information on the early stages of the folding.

Šponer and co-workers have recently performed a series of molecular dynamics (MD) simulations probing the stability of various possible intermediates on a microsecond time scale and providing an atomic-detailed insight into the early stage of the folding.^[404–407] These studies suggest that the distribution of *syn/anti* glycosidic angles in the final structure determines the relative movements allowed between the strands during the refolding process. In particular, a 2-tetrads quadruplex, proposed as possible intermediate on the basis of NMR data by Zhang *et al.*,^[119] can lead to the formation of a 3-tetrad quadruplex by slippage of one strand only in an all-*anti* conformation, while a strand-slippage with *syn/anti* glycosidic angles is not favorable as it may lead to steric clashes (**Figure 44**). As an intermediate in the G-quadruplex folding pathway, a G-triplex could potentially be formed by binding/unbinding of one strand. MD simulations show, however, that a G-triplex with all-parallel strands and all-*anti* guanine residues is not stable on a microsecond time-scale (**Figure 44**). The existence of the G-triplex has been a matter of some debate. Stand-alone G-triplex has not been observed until now for three G-stretches sequences derived from human telomeric DNA, nevertheless optical tweezers studies, MD and CD investigations suggested the G-triplex structure as a plausible intermediate.^[400,405,408] However, it is still not clear if these structures exist, if they are sufficiently stable to be considered as an intermediate and what is their exact structure.

On the other hand, the G-hairpin has been also indicated as potential intermediate on the G-quadruplex folding pathway. Mashimo *et al.* showed with molecular orbital (MO) calculation that the GG Hoogsteen base pair has a stability comparable to the one of AT Watson-Crick base pair and proposed the hairpin formation as first intermediate to the formation of G-quadruplex structures.^[409] Images of G-hairpins and G-triplexes in DNA origami frames have been captured in

2014 by Rajendran *et al.* using high-speed atomic force microscopy (HS-AFM).^[410] The formation of an hairpin as a possible intermediate is also supported by a recent NMR study from Plavec group. The NMR investigation on the folding pathway of the bimolecular quadruplex formed by *Oxytricha nova* telomeric DNA (G₄T₄G₄) suggested the presence of symmetric bimolecular intermediates where all the guanines are involved in GG N1-carboyl symmetric base pairs.^[411]

It has to be noted that while the structural features of the various species involved in the folding can be unambiguously assigned only using atomic-resolution methods, the number of species involved in the folding and their distribution of population can be assessed by single-molecule Förster resonance energy transfer (smFRET).^[412,413]

Other interesting details on the dynamic of the folding can be provided by MD simulations.^[407] For example, MD reveals that the propeller loops are in general very strained and their formation constitutes a peculiar aspect of the folding pathway, in agreement with our hypothesis discussed in Chapter 7.1. Moreover, MD allows to identify external, non-channel cation binding sites with high degree of cation delocalization which cannot be detected by X-ray and may be useful to promote the folding (*i.e.*, narrow edgewise loops have been shown to potentially coordinate a K⁺ cation).^[117]

A combined NMR/MD approach could possibly be used in the future to probe the folding on different time-scales.

7.1 Research article: Involvement of long-lived intermediate states in the complex folding pathway of the human telomeric G-quadruplex

Bessi I., Jonker H. R. A., Richter C., Schwalbe H.

Angewandte Chemie International Edition, 2015, 13; 54(29): 8444-8.

In this article, the K^+ -induced folding kinetic of the human telomeric sequence d[TTGGG(TTAGGG)₃A] (Tel24) was investigated using real-time NMR. Analysis of the NMR-derived kinetic traces provided atomistic insight into the folding mechanism of telomeric G-quadruplex.

After injection of K^+ , two distinct folded states were detected: a major conformation (hybrid-1) and a previously not characterized minor conformation (hybrid-2), which is formed faster than the more stable hybrid-1. Interestingly, a partially unfolded state was also populated during the folding. We propose that the thermodynamic (hybrid-1) and kinetic (hybrid-2) conformations equilibrate slowly via the partially unfolded intermediate state, which can be described as an ensemble of hairpin-like structures.

All the results presented in this work were collected and analysed by the author of this thesis.

Involvement of Long-Lived Intermediate States in the Complex Folding Pathway of the Human Telomeric G-Quadruplex**

Irene Bessi, Hendrik R. A. Jonker, Christian Richter, and Harald Schwalbe*

Abstract: The energy landscapes of human telomeric G-quadruplexes are complex, and their folding pathways have remained largely unexplored. By using real-time NMR spectroscopy, we investigated the K^+ -induced folding of the human telomeric DNA sequence 5'-TTGGG(TTAGGG)₃-3'. Three long-lived states were detected during folding: a major conformation (hybrid-1), a previously structurally uncharacterized minor conformation (hybrid-2), and a partially unfolded state. The minor hybrid-2 conformation is formed faster than the more stable hybrid-1 conformation. Equilibration of the two states is slow and proceeds via a partially unfolded intermediate state, which can be described as an ensemble of hairpin-like structures.

In the presence of monovalent cations, such as K^+ or Na^+ , G-rich DNA sequences can form a G-quadruplex structure by the stacking of multiple G-quartets, each composed of four guanine residues arranged cyclically and stabilized by Hoogsteen hydrogen bonds (Figure 1a).^[1,2] G-quadruplexes are important in biology^[3] but also utilized in various nanotechnological applications.^[4] An overwhelming wealth of data show human telomeric DNA polymorphism (Figure 1b,c). The dominant conformation is determined mainly by the presence of flanking nucleotides and the nature of the stabilizing cation, but also by molecular crowding.^[5] Owing to their appearance at the ends of telomeres, G-quadruplexes are proposed targets for pharmacological intervention.^[6,7] Moreover, their ability to adopt substantially different conformations challenges our understanding of protein–quadruplex interactions and their inhibition by drugs.^[8]

We herein describe the application of time-resolved NMR spectroscopy to investigate the K^+ -induced folding of the human telomeric sequence Tel24, a well-studied system that adopts the hybrid-1 conformation (Figure 1d, left).^[9] We show that after the injection of K^+ ions, the folding of Tel24

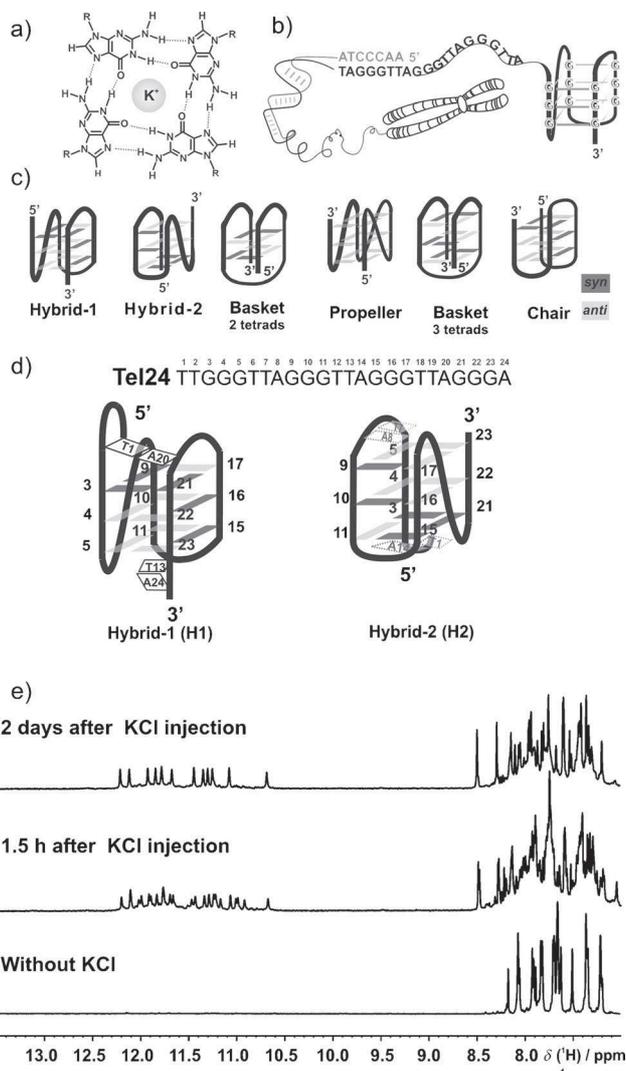


Figure 1. a) General structure of a G-quartet. b) The 3' single-stranded overhang of human telomeric DNA adopts a G-quadruplex fold. c) Polymorphism of human telomeric DNA: hybrid-1,^[9–11] hybrid-2,^[11,12] two-tetrad basket,^[13,14] propeller,^[15,16] three-tetrad basket,^[17] chair.^[18] d) Sequence and numbering of Tel24 DNA. Left: Folding topology of the major conformation of Tel24 (hybrid-1) as determined by NMR spectroscopy by Luu et al.^[9] Right: Folding topology of the kinetically favored conformation of Tel24 (hybrid-2) as determined in this study, with proposed capping structures indicated by gray dashed lines. Guanine residues with an *anti* orientation are colored light gray, whereas *syn*-oriented guanine residues are colored dark gray. e) 1D ¹H NMR spectra of Tel24 at different stages of folding induced by the addition of K^+ ions. Experimental conditions: 0.2 mM DNA, 25 mM Bis-Tris buffer, pH 7.0, 10% D₂O/90% H₂O, 298 K, 600 MHz.

[*] I. Bessi, Dr. H. R. A. Jonker, Dr. C. Richter, Prof. Dr. H. Schwalbe
Institute for Organic Chemistry and Chemical Biology
Center of Biomolecular Magnetic Resonance (BMRZ)
Goethe University Frankfurt/Main
Institution Max-von-Laue-Strasse 7, 60438 Frankfurt (Germany)
E-mail: schwalbe@nmr.uni-frankfurt.de
Homepage: <http://schwalbe.org.chemie.uni-frankfurt.de>

[**] This research was supported by the State of Hesse (LOEWE project: Synchembio and BMRZ) and by the DFG (Cluster of Excellence: Macromolecular Complexes). We thank Elke Stirnal for the HPLC purification and Dr. Boris Fürtig, Dr. Kai Schlepckow, and Dr. Anna Lena Lieblein for stimulating discussions.

Supporting information for this article, including experimental details, is available on the WWW under <http://dx.doi.org/10.1002/anie.201502286>.

undergoes kinetic partitioning, and three conformations are populated: hybrid-1 (H1), hybrid-2 (H2), and an unfolded conformation (U) that is not stabilized by hydrogen bonds. The formation of the less stable conformation H2 is kinetically favored. Refolding of the H2 conformation involves extensive reorganization of the G-quadruplex to the more stable H1 conformation via partially structured states and proceeds on a very slow time scale.

The K^+ -induced folding and subsequent refolding of the Tel24 G-quadruplex requires several hours at 298 K. Two distinct G-quadruplex conformations were detected 1.5 h after the injection of K^+ ions (Figure 1 e). Whereas the major conformation adopts the H1 fold, we characterized the folding topology of its minor conformation, H2, which is formed more rapidly. The imino and aromatic guanine hydrogen-atom resonances of both conformations were assigned (Figure 2; see also Figure S5 in the Supporting Information), and the intratetrad 1H_1 - 1H_8 connectivities were derived (see Figure S7b for atomic numbering). H-D exchange experiments showed that the imino hydrogen atoms of the inner G-quartet exchange much slower than those on the external quartets.^[19] After solvent exchange, the imino hydrogen-atom signals of the inner quartet (residues G22, G16, G10, and G4) and of residue G15 in the H2 conformation were still visible (see Figure S6), and the signal from residue G15, which defines an outer quartet, disappeared more quickly than the others.

At room temperature, the population of the minor conformation H2 is not sufficiently long-lived to be characterized by NMR spectroscopy. NMR spectroscopic refolding experiments (see Figures S7–10 for the results of 1H , 1H NOESY and 1H , ^{13}C HSQC experiments) were therefore conducted at 288 K with a DNA concentration of less than 1 mM to avoid aggregation. Assignment of the major conformation H1 was performed on the basis of previously

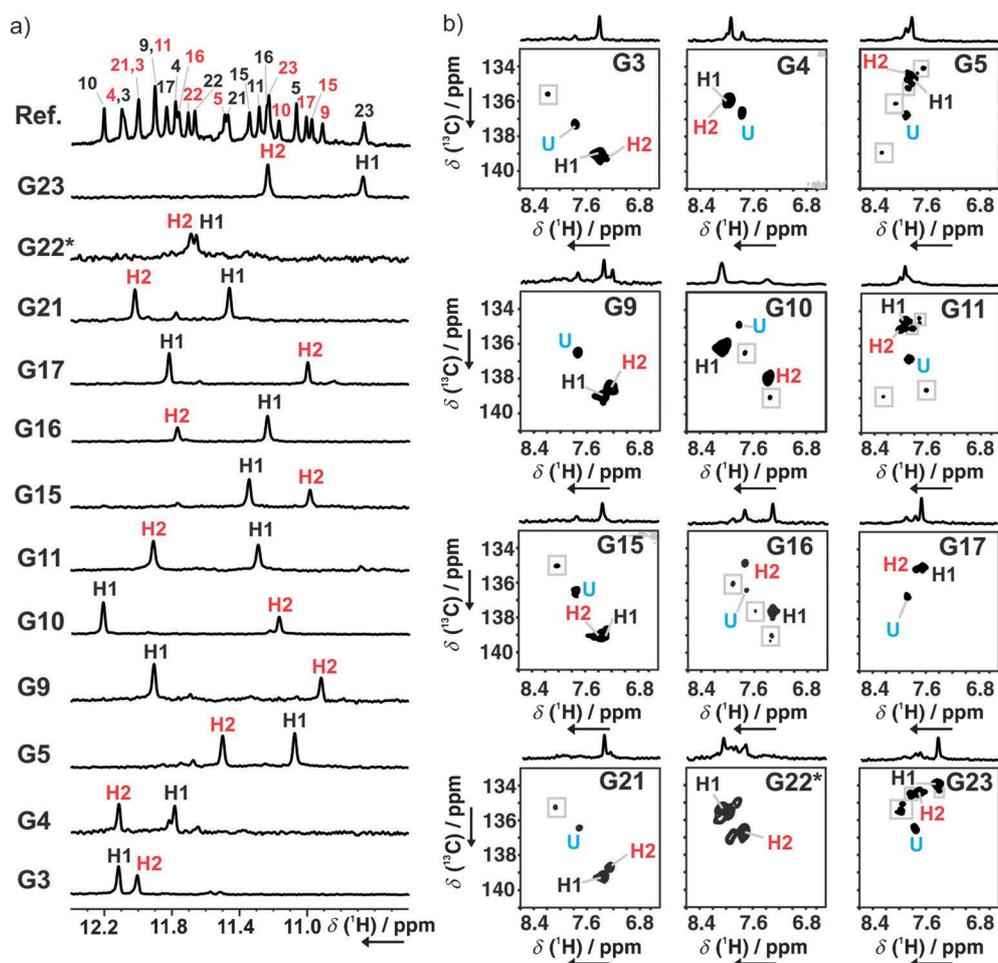


Figure 2. a) NMR spectroscopic assignment of guanine imino hydrogen atoms in the major (H1) and minor (H2) conformation of Tel24 by the use of 1D 1H , ^{15}N -filtered experiments on site-specifically labeled samples. The reference spectrum at the top is annotated with the complete signal assignment. b) NMR spectroscopic assignment of aromatic guanine hydrogen atoms by the use of 2D 1H , ^{13}C HSQC experiments on site-specifically labeled samples. Label color code: black for the hybrid-1 conformation (H1), red for the hybrid-2 conformation (H2), and cyan for the unfolded state (U). Additional signals arising from unassigned minor species are framed with gray boxes. NMR spectra were recorded after the induction of folding ($[KCl]:[DNA] = 140:1$) at room temperature and cooling down to 288 K of samples with uniformly ^{13}C and ^{15}N site-specifically labeled oligonucleotides at the indicated position. Experimental conditions: 100–250 μM DNA, 25 mM Bis-Tris buffer, pH 7.0, 10% $D_2O/90\%$ H_2O , 288 K, 600 MHz. *The DNA labeled at G22 contained impurities.

reported data.^[9] With the unambiguous assignment of exchangeable and non-exchangeable guanine hydrogen atoms of the minor conformation by the use of selectively labeled samples (Figure 2), we could assign the 1H_1 - 1H_8 region of the NOESY spectrum and determine the quartet arrangement typical of the H2 conformation (Figure 1 d, right; see also Figure S7).^[11,12] All folding and refolding kinetics were determined at 298 K in real-time NMR spectroscopic experiments (see the Supporting Information for experimental details). KCl was manually injected at 298 K into the NMR tube to give a $[K^+]/[DNA]$ ratio of 140:1. Kinetic traces for different residues were extracted from the same experiment on a sample at natural abundance. Only well-resolved imino peaks (see Figure S12) were chosen for the kinetic analysis, which excluded the analysis of signals

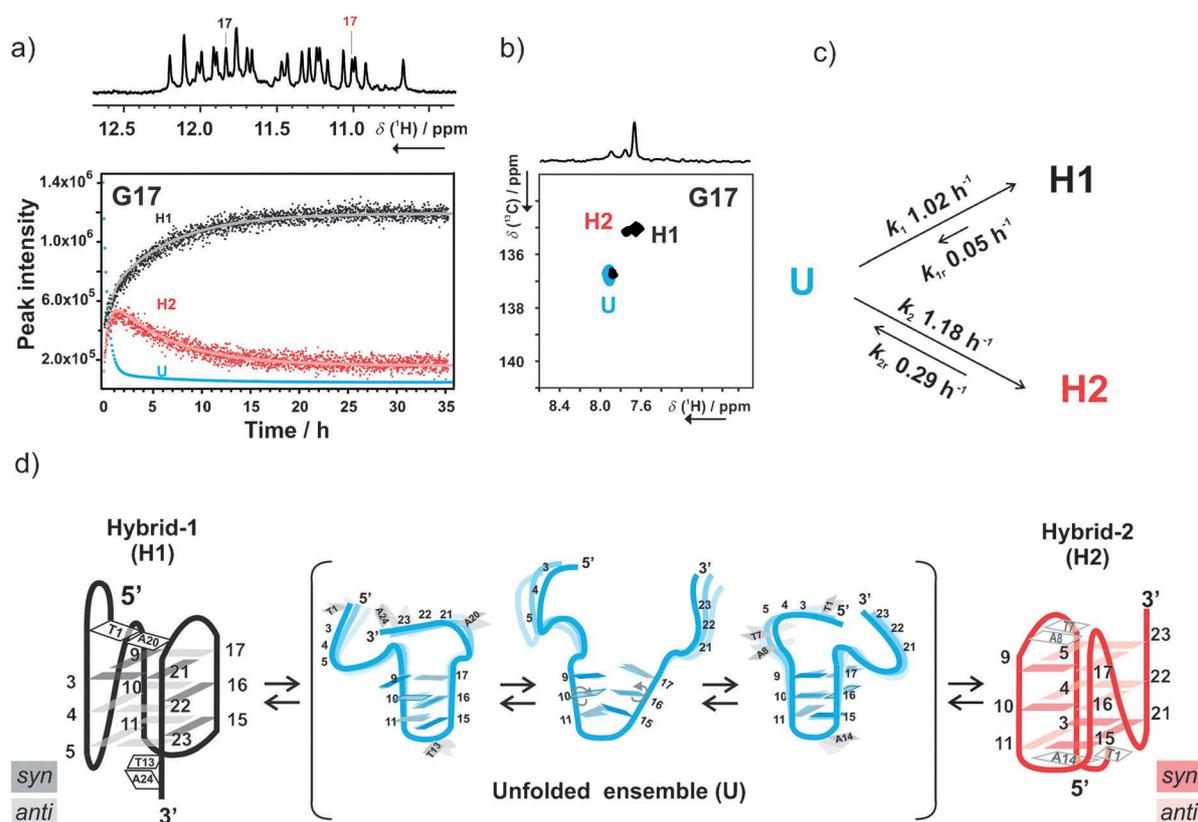


Figure 3. a) Kinetic traces describing the intensity of the imino peaks of residue G17 as a function of time at 298 K. Imino signals of G17 in the hybrid-1 (H1, black) and hybrid-2 conformation (H2, red) are marked on the 1D ^1H spectrum recorded 2.5 h after inducing folding at 298 K. b) Aromatic region of the ^1H , ^{13}C HSQC spectrum of the sample selectively labeled at G17 before the induction of folding (cyan) and 3 days after folding induction (black) at 298 K. Label color code: black for the H1 conformation,^[9] red for the H2 conformation, and cyan for the unfolded state (U). c) Kinetic model proposed to fit the data. Average rate constants obtained from the global fitting of well-resolved imino peaks (see Figure S12 and Table S1) are shown. Fitting curves are displayed in (a) as solid lines and result from the global fitting of the kinetic traces with the proposed mechanism (see the Supporting Information for details). d) Folding topologies involved in the proposed folding mechanism for Tel24. Proposed capping structures for H2 are in gray.

originating from either aromatic sites or thymine methyl groups.

Kinetic plots of the G17 imino hydrogen-atom signal intensities for the H1 (black) and the H2 (red) conformations are shown as an example in Figure 3a. ^1H , ^{13}C HSQC spectra of several site specifically labeled samples revealed the presence of a residual amount of an unfolded state (Figure 3b), which remained even months after the induction of folding and storage of the samples at room temperature. The signal of the unfolded state could be detected even at higher K^+ concentrations (see Figure S11).

From the analysis of folding rates (see Table S1 in the Supporting Information), we propose a folding model for Tel24 (Figure 3c) that involves three states: H1 (black), H2 (red), and an unfolded state U (cyan). At $T=298$ K, the H1 and H2 conformations are formed with average rate constants k_1 and k_2 equal to (1.02 ± 0.02) and $(1.18 \pm 0.03) \text{ h}^{-1}$, respectively. The re-equilibration of the two conformers is slow and proceeds through the formation of a state U without detectable hydrogen bonds, at least at neutral pH. The unfolding of the minor conformation H2 is one order of magnitude faster than the unfolding of H1 ($k_{1r} = (0.049 \pm 0.001) \text{ h}^{-1}$ and $k_{2r} = (0.293 \pm 0.006) \text{ h}^{-1}$). Whereas in the initial

folding phase at room temperature both folded conformations are populated to a comparable extent, at higher temperatures the H2 population is larger than that of H1 (see Figure S13b). At 310 K, 6 min after the injection of KCl , the population of H2 was approximately twice as large as the population of H1.

The effect of temperature on the rate constants was investigated in the range from 298 to 316 K, in which the kinetic traces showed a biphasic behavior. Kinetic traces of the imino hydrogen atom belonging to residue G5 were fitted with the mechanism shown in Figure 3c (see Table S2), and activation parameters shown in Table 1 were derived from Eyring analysis (see Figure S13) of the rate constants.

$$\ln \frac{k}{T} = \ln \frac{k_B}{h} + \frac{\Delta S^\ddagger}{R} - \frac{\Delta H^\ddagger}{RT} \quad (1)$$

The conformation of the two folded states can clearly be assigned to hybrid structures with a very distinct overall topology; the conformation of the unfolded state, however, is less clear. We tentatively describe this state as an ensemble of prefolded hairpin structures in which the inner G-stretches are partially formed and the 3' and 5' ends are unfolded

Table 1: Activation parameters resulting from the linear regression of the kinetic rates to Equation (1).^[a] Errors result from the fitting.

	k_1	k_2	k_{1r}	k_{2r}
ΔH^\ddagger [kJ mol ⁻¹]	191.7 ± 3 %	226.4 ± 6 %	131.1 ± 10 %	135.8 ± 10 %
ΔS^\ddagger [J K ⁻¹ mol ⁻¹]	397.3 ± 3 %	514.8 ± 6 %	170.3 ± 10 %	201.4 ± 10 %
E_a (at 298 K) [kJ mol ⁻¹]	194.2 ± 3 %	228.9 ± 6 %	133.6 ± 10 %	138.2 ± 10 %

[a] In Equation (1), k_B is the Boltzmann constant, h is the Planck constant, and R is the gas constant (see the Supporting Information for values).

(Figure 3d). In support of this hypothesis, we detected very broad signal(s) between 10.5 and 11.0 ppm at pH 7 and 298 K (see Figure S2b). According to Čeru et al., these signals could be due to the presence of preorganized structures with GG N1-carbonyl symmetric base pairs.^[20] However, even in 100 % H₂O (and with an insert filled with D₂O to obtain an NMR lock signal), 1D ¹H,¹⁵N-edited NMR spectra recorded on ¹³C,¹⁵N site specifically labeled Tel24 at pH 7.0 did not show any clear signal at around 11 ppm, in which region the imino resonances of the GG N1-carbonyl symmetric base pairs are expected.^[20] We did not conduct experiments at more acidic pH values.

The conformational switch of different unfolded hairpin conformations that leads to the formation of either the H1 or H2 folded state requires a coordinated swap of the nucleobases belonging to the inner G-stretches, together with an *syn/anti* transition of G10–G16 (Figure 3d). Since Mg²⁺ was reported to have a stabilizing effect on hairpin structures,^[21,22] we tested whether the addition of Mg²⁺ to a K⁺-free solution of 0.3 mM Tel24 in Bis-Tris buffer at pH 7.0 stabilized prefolded hairpin-like structures (see Figures S3 and S4); however, at concentrations up to 4 mM Mg²⁺, we did not observe any signals that would indicate stable-hairpin formation.

Over the years, different hypotheses on the conformation of species involved in the folding process of telomeric G-quadruplexes have been suggested. On the basis of circular dichroism (CD), intrinsic 2-aminopurine fluorescence, and fluorescence resonance energy transfer (FRET), Gray et al. proposed that G-quadruplex folding is multiphasic, with three different intermediates formed sequentially along the folding pathway.^[8] According to their analysis, after the addition of K⁺ cations, prefolded hairpins rapidly collapse into an antiparallel quadruplex (e.g. a chair), which in turn evolves into a third intermediate (hypothesized to be a triplex), and finally forms the hybrid structure. However, the nature of these partially unfolded species and short-lived folding intermediates has remained controversial, especially because the disentanglement of fluorescence and CD signatures (see Figure S1) due to various quadruplex conformations is difficult. The existence of G-triplexes along the folding pathway of the telomeric G-quadruplex has been an object of debate in recent years.^[8,23–26] The G-hairpin has also been indicated as a potential intermediate on the G-quadruplex

folding pathway.^[18,22,27] Moreover, NMR spectroscopic investigation of the folding pathway of the bimolecular quadruplex formed by *Oxytricha nova* telomeric DNA (G₄T₄G₄) suggested the presence of symmetrical bimolecular intermediates in which all guanine residues are involved in GG N1-carbonyl symmetric base pairs.^[20]

Our data indicate that neither a chair nor a triplex is a long-lived intermediate. Furthermore, previous kinetic studies have failed to delineate the parallel folding pathway and the coexistence of two hybrid conformations together with residual unfolded species, presumably owing to the insensitivity of the spectroscopic probes used. We observed only two long-lived species with imino hydrogen atoms protected from exchange with water: H1 and H2. Although there might be other transient species involved in the formation of the H1 conformation, they are neither sufficiently long-lived nor sufficiently populated to be characterized by NMR spectroscopy. A more appropriate method to investigate these short-lived marginally stable intermediates could be molecular-dynamics (MD) simulations.^[28]

Analysis of the spectra of site specifically labeled samples (Figure 2A; see also Figure S5) indeed reveals the presence of more than two imino signals for most of the residues in the early stage of folding. The population of these species is less than 5 % and is further reduced beyond detection few hours after the induction of folding. These additional species could be either on-pathway or off-pathway intermediates. According to the number of imino signals observed in the 2D ¹H,¹⁵N SOFAST-HMQC spectrum, the residues belonging to the inner quartet (G22, G10, G16, and G4) are less affected by structural heterogeneity than the guanine residues defining the external tetrads. Therefore, we speculate that these additional short-lived intermediates conserve the general hybrid scaffold but present different or partially formed capping structures. This interpretation is in line with the observation that AT base pairs and triplets within the loops play a crucial role in G-quadruplex stabilization. In fact, Dai and co-workers suggested previously on the basis of the comparison of 1D ¹H imino patterns that sequences derived from human telomeric DNA are always in equilibrium between the two forms H1 and H2, whereby the population ratio is determined by the nature of the 3' and 5' flanking nucleotides.^[5,12,14] Our 2D spectroscopic data, however, unambiguously demonstrate that the minor conformation is the H2 fold, and that it is populated during the K⁺-induced folding of H1. Furthermore, we could show that the formation of H2 is kinetically favored and that the interconversion of the two hybrid forms proceeds via an unfolded state, whose substantial importance was previously remarked on only by Lee et al.^[29]

It is interesting to note that the kinetics of i-motif formation of the C-rich complementary sequence of human telomeric G-quadruplex also follows kinetic partitioning.^[30] Still, the question of why the H2 (minor) conformation is formed faster than the H1 (major) conformation remains to be discussed. NOESY spectra indicate that the stabilizing structures of H2 (Figure 3d; see also Figure S10) can be formed from the hairpin ensemble by swinging back of the 5'-end G-stretch. On the other hand, the stabilizing caps of H1

(A24:T13 and T1:A20 base pairs) can be formed from the hairpin ensemble only after formation of the T18–T19–A20 edgewise loop and of the T6–T7–A8 propeller loop, which requires not only that the 3' end swings back from the hairpin, but also the inversion of strand orientation. This extensive reorganization might slow down the formation of H1. The equilibrium population ratio of the two conformers can be explained in terms of the possibility of forming specific stabilizing capping structures at the 3' and 5' ends. As reported previously, the H2 structure is stabilized by a specific T:A:T cap at the 3' end in different G-quadruplexes, which cannot be formed by Tel24.^[11,12] This hypothesis was tested by replacing residue 24 A for T. This exchange indeed resulted in a significant increase in the population of the minor conformation (see Figures S14 and S16).

Analysis of the kinetic traces of other sequences derived from human telomeric DNA (see Figures S14 and S15) reveals that, irrespective of the sequence of the flanking nucleotides, the folding process is biphasic and proceeds through the initial formation of a kinetic product that slowly equilibrates with a thermodynamic product. This finding suggests that kinetic partitioning might be a common property of G-quadruplexes, and presumably linked to their polymorphism.

In conclusion, we propose for Tel24 a folding and refolding mechanism in which the conformational equilibration between the two hybrid forms is slow and proceeds via an “unfolded” state. Although we are aware that the elongated telomeric sequence might behave differently in vivo,^[31] this study contributes to efforts to elucidate the complex energy landscape of G-quadruplex folding. For the first time it provides an atomistic description of the species involved in the process of folding and refolding of the human telomeric G-quadruplex. This study points out that the presence of partially unfolded regions might be an intrinsic property of human telomeric DNA and necessary to modulate the binding of telomere-binding proteins, such as POT1 and telomerase enzyme, to the single-stranded telomeric overhang. The elongation rate reported for human telomerase enzyme is 1 nts⁻¹;^[32] therefore, on a biological time scale kinetic and thermodynamic conformers are both present and coexist with partially unfolded regions. The human telomeric G-quadruplex, with its diversity of structures, is a challenging target for medicinal chemistry, and the development of small molecules that act as selective stabilizers for the thermodynamic conformation might not be the most effective therapeutic strategy, given that the kinetically favored conformation is long-lived, as shown herein. The findings will therefore impact the development of drugs that bind to telomeric DNA and the use of G-quadruplexes in nanobiotechnology.

Keywords: conformation analysis · DNA structures · folding kinetics · G-quadruplexes · real-time NMR spectroscopy

How to cite: *Angew. Chem. Int. Ed.* **2015**, *54*, 8444–8448
Angew. Chem. **2015**, *127*, 8564–8568

- [1] M. Gellert, M. Lipsett, D. Davies, *Proc. Natl. Acad. Sci. USA* **1962**, *48*, 2013–2018.
- [2] D. Sen, W. Gilbert, *Nature* **1988**, *334*, 364–366.
- [3] P. Murat, S. Balasubramanian, *Curr. Opin. Genet. Dev.* **2014**, *25*, 22–29.
- [4] L. A. Yatsunyk, O. Mendoza, J.-L. Mergny, *Acc. Chem. Res.* **2014**, *47*, 1836–1844.
- [5] J. X. Dai, M. Carver, D. Z. Yang, *Biochimie* **2008**, *90*, 1172–1183.
- [6] S. Neidle, *FEBS J.* **2010**, *277*, 1118–1125.
- [7] S. A. Ohnmacht, S. Neidle, *Bioorg. Med. Chem. Lett.* **2014**, *24*, 2602–2612.
- [8] R. D. Gray, J. O. Trent, J. B. Chaires, *J. Mol. Biol.* **2014**, *426*, 1629–1650.
- [9] K. N. Luu, A. T. Phan, V. Kuryavyi, L. Lacroix, D. J. Patel, *J. Am. Chem. Soc.* **2006**, *128*, 9963–9970.
- [10] J. X. Dai, C. Punchihewa, A. Ambrus, D. Chen, R. A. Jones, D. Z. Yang, *Nucleic Acids Res.* **2007**, *35*, 2440–2450.
- [11] A. T. Phan, V. Kuryavyi, K. N. Luu, D. J. Patel, *Nucleic Acids Res.* **2007**, *35*, 6517–6525.
- [12] J. X. Dai, M. Carver, C. Punchihewa, R. A. Jones, D. Z. Yang, *Nucleic Acids Res.* **2007**, *35*, 4927–4940.
- [13] K. W. Lim, S. Amrane, S. Bouaziz, W. Xu, Y. Mu, D. J. Patel, K. N. Luu, A. T. Phan, *J. Am. Chem. Soc.* **2009**, *131*, 4301–4309.
- [14] Z. Zhang, J. Dai, E. Veliath, R. A. Jones, D. Yang, *Nucleic Acids Res.* **2010**, *38*, 1009–1021.
- [15] G. N. Parkinson, M. P. H. Lee, S. Neidle, *Nature* **2002**, *417*, 876–880.
- [16] B. Heddi, A. T. Phan, *J. Am. Chem. Soc.* **2011**, *133*, 9824–9833.
- [17] Y. Wang, D. J. Patel, *Structure* **1993**, *1*, 263–282.
- [18] T. Mashimo, H. Sugiyama, *Nucleic Acids Symp. Ser.* **2007**, 239–240.
- [19] F. W. Smith, J. Feigon, *Nature* **1992**, *356*, 164–168.
- [20] S. Čeru, P. Šket, I. Prislán, J. Lah, J. Plavec, *Angew. Chem. Int. Ed.* **2014**, *53*, 4881–4884; *Angew. Chem.* **2014**, *126*, 4981–4984.
- [21] A. Bugaut, P. Murat, S. Balasubramanian, *J. Am. Chem. Soc.* **2012**, *134*, 19953–19956.
- [22] A. Rajendran, M. Endo, K. Hidaka, H. Sugiyama, *Angew. Chem. Int. Ed.* **2014**, *53*, 4107–4112; *Angew. Chem.* **2014**, *126*, 4191–4196.
- [23] D. Koirala, T. Mashimo, Y. Sannohe, Z. Yu, H. Mao, H. Sugiyama, *Chem. Commun.* **2012**, *48*, 2006–2008.
- [24] P. Stadlbauer, L. Trantírek, T. E. Cheatham, J. Koča, J. Sponer, *Biochimie* **2014**, *105*, 22–35.
- [25] M. Bončina, J. Lah, I. Prislán, G. Vesnaver, *J. Am. Chem. Soc.* **2012**, *134*, 9657–9663.
- [26] V. Limongelli, S. De Tito, L. Cerofolini, M. Fragai, B. Pagano, R. Trotta, S. Cosconati, L. Marinelli, E. Novellino, I. Bertini, A. Randazzo, C. Luchinat, M. Parrinello, *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 2269–2273; *Angew. Chem.* **2013**, *125*, 2325–2329.
- [27] T. Mashimo, H. Yagi, Y. Sannohe, A. Rajendran, H. Sugiyama, *J. Am. Chem. Soc.* **2010**, *132*, 14910–14918.
- [28] P. Stadlbauer, M. Krepl, T. E. Cheatham, J. Koca, J. Sponer, *Nucleic Acids Res.* **2013**, *41*, 7128–7143.
- [29] J. Y. Lee, B. Okumus, D. S. Kim, T. Ha, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 18938–18943.
- [30] A. L. Lieblein, J. Buck, K. Schlepckow, B. Fürtig, H. Schwalbe, *Angew. Chem. Int. Ed.* **2012**, *51*, 250–253; *Angew. Chem.* **2012**, *124*, 255–259.
- [31] R. Hänsel, F. Löhr, L. Trantírek, V. Dötsch, *J. Am. Chem. Soc.* **2013**, *135*, 2816–2824.
- [32] H. Hwang, P. Opreško, S. Myong, *Sci. Rep.* **2014**, *4*, 6391.

Received: March 11, 2015

Published online: June 2, 2015

8 G-quadruplex in nanotechnology

Beyond encoding the genetic information and directing the replication of the genetic material, DNA self-assembly can be used to control the structure of matter and build nanodevices.^[414–416]

Due to its intrinsic nano dimensions and to its self-assembly properties, DNA is an attractive tool for structural nanotechnology to “put what you want where you want it in three dimensions (3D), when you want it there”.^[415]

Around 1996, two different approaches were developed to obtain DNA-based rigid building blocks for programmable nanomaterials. The first approach, used in the labs of Seeman, was based on the careful design of branched tiles resulting from hybridization and intertwining of DNA strands. These tiles were then used to build 2D and 3D nano arrays. The so-called DNA origami approach is also based on hybridization and intertwining of strands, but uses a plasmid DNA scaffold, which is folded in single nanostructures with the help of staple oligonucleotides. The second method, introduced in Mirkin’s lab, relies on the functionalization of a rigid core (usually a gold nanoparticle), with single stranded DNA that can template the particle-particle assembly to form long-range ordered 3D lattices and even single crystals.^[417]

Although the classic duplex B-conformation based on the Watson-Crick interaction scheme has by far been the most commonly used “tool” in DNA-based nanotechnology, non-canonical base pairing schemes and alternative DNA structures offer additional possibilities for building DNA-based nanomaterials. As shown in Chapter 1, DNA molecules can self-assemble *via* many different base-pairing schemes and assume a variety of conformations having some advantages over the classical B-DNA duplex, namely the capability to respond to chemical stimuli (H^+ , metal cations) and enhanced thermal stability.^[6] Furthermore, the possibility of different hydrogen-bonding patterns results in an extensive polymorphism: the conformational transition between different structures (such as duplex/quadruplex, B/Z, i-motif/single strand) can be triggered with various external stimuli and used as engine to drive switchable nanomachines.

The sketch depicted in **Figure 45** summarizes a number of possible G-quadruplex-based applications in the different branches of nanotechnology.

In the field of DNA-based nanoarchitectures, G-quadruplex structures have been used one decade ago as “molecular glue” to control the self-assembly of Au nanoparticles *via* cations.^[418] Heckel and co-workers showed that G-quadruplex can also be used to glue DNA mini-circles (**Figure 45a**), and in principle larger DNA objects that can be monitored by atomic force microscopy (AFM).^[419] In the group of Sugiyama, two DNA sequences able to form inter-molecular G-quadruplexes have been embedded in a DNA origami frame and the formation of the G-quadruplex has been visualized by AFM (**Figure 45b**). The same technique has been used to characterize G-hairpin and G-triplex structures which are possibly formed during the G-quadruplex folding (see Chapter 7 for more details).^[410]

There are also reports on applications where G-quadruplexes have been used to control and modify surface properties. For example, nano-pores containing guest molecules were

functionalized with C-rich oligonucleotides and blocked with Au nanoparticles linked to G-rich oligonucleotides able to hybridize with the C-rich counterpart. Addition of K^+ cations triggers the formation of G-quadruplexes and the release of the guest molecules (**Figure 45c**).^[420] Furthermore, the thrombin binding aptamer (TBA) G-quadruplex, able to bind with high affinity and specificity to the human blood clotting factor α -thrombin and acting therefore as anti-coagulant, has been encoded into DNA tiles forming a 2D-lattice (**Figure 45d**). The result is a regular spatial distribution of TBA aptamers, which can be used as a sensing array with TBA acting as anchor able to bind a specific probe (thrombin).

G-quadruplexes in combination with i-motif structures have been exploited as functional material for computing. Using pH and cations as input, these systems have been shown to behave as logic gates with potential applications in DNA computing (**Figure 45e**).^[421]

Willner's group has developed a variety of supramolecular architectures able to sense analytes and induce the formation of the G-quadruplex Horseradish Peroxidase (HRP) mimicking DNAzyme as catalytic label (**Figure 45f**). Binding of the G-quadruplex DNAzyme to hemin catalyses the reduction of H_2O_2 and the oxidation of a target molecule that can be detected with chemiluminescent or colorimetric methods. Careful design of the multicomponent supramolecular assembly containing the G-quadruplex DNAzyme allows to detect different analytes (metal ions, cocaine, ATP, thrombin, target DNA).^[422]

Another growing field of the DNA-based nanotechnology is the development of nanomaterials able to respond to external stimuli in a dynamic way: the cyclic alternation between two different DNA conformations is the basic idea of many nano-switches. A DNA switch must be triggered by an appropriate "fuel" and must be reconfigurable in the initial state.

Nanomotors based on a G-quadruplex/duplex equilibrium use a DNA strand replacement mechanism to trigger the conformational switch. The G-quadruplex is disrupted by hybridization with the complementary C-strand (fuel). Using a third strand (anti-fuel) with appropriate sticky-ends, the G-quadruplex is regenerated *via* displacement of the C-rich strand and production of a "waste" duplex. The TBA G-quadruplex has also been embedded in such a molecular motor to control the release of thrombin (**Figure 45g**).^[6] Cations have also been used as chemical triggers. For example, a partially complementary duplex containing a G-rich internal loop was labeled with a ferrocene and deposited on a gold electrode. Upon addition of K^+ (fuel), the G-rich loops can fold into a G-quadruplex structure and the duplex can be regenerated by addition of a crown ether able to sequester the K^+ cation (**Figure 45h**). The cyclic switch between the two different conformations produces a substantial variation in the ferrocene-gold electrode distance that can be monitored by voltammetry (see reference [423] and references given therein).

Instead of a chemical stimulus, light can also be used as triggering input. Light allows a very precise temporal and spatial control on the nanoswitch and does not produce any "waste" (e.g., a fuel/anti-fuel duplex DNA). The ideal switch upon irradiation should produce with high yield a long-lived state with a change in end-to-end distance, sufficient to trigger the self-assembly of nanomaterials or affect the biological activity of the target. For *in vivo* biological applications it should be non-toxic, stable and able to switch at a wavelength long enough to be not harmful for

cell and tissues. Stilbenes, spiropyranes and diarylethenes have been used to control the switch of nucleic acids, however the azobenzene group (Az) is by far the most popular photoswitch element used because of the high efficiency and robustness of its E/Z photoisomerization.^[424–426] Woolley and co-workers have recently investigated how to tune the photochemical properties of azobenzene derivatives to make them compatible with biological applications.^[427] As discussed in Chapter 4, G-quadruplexes are involved in regulation of biological processes and the inclusion of an appropriate photoswitch could be used to reversibly photo-control gene expression *in cell*.

The photocontrol of G-quadruplexes with Az (red color code in **Figure 45i-k**) has been achieved 1) using Az as photoswitchable ligand or 2) incorporating Az directly in the oligonucleotide by solid phase synthesis.

Zhou has used strategy 1) to prepare an Az derivative that in *E*-conformation is able to template the formation of G-quadruplex in absence of cations, while in *Z*-conformation induces the dissociation of the quadruplex (**Figure 45i**). Tan and co-workers adopted approach 2) to photo-control blood clotting. The TBA regulatory domain was connected *via* a linker to an inhibitory domain composed of the TBA complementary sequence functionalized with Az groups. Each Az group was tethered to an acyclic, non-ribose linker inserted in the phosphate backbone. The TBA aptamer is free to interact with the thrombin and to inhibit coagulation with the Az groups in the *E*-state, while upon irradiation with visible light, the regulatory domain hybridizes to the inhibitory domain forming a duplex stabilized by Az intercalating between the Watson-Crick base pairs (**Figure 45j**).

Following the strategy reported by Wu *et al.* for the development of an Az-capped DNA hairpin,^[428] Heckel and co-workers have inserted an Az unit directly in the phosphate backbone between two G-rich strands, replacing ideally a possible loop in a bimolecular G-quadruplex. The Az-derivative is able to form a well-defined intermolecular G-quadruplex structure, whose structural characteristics and photoswitching behavior will be described in the next Chapter. Such G-quadruplex-based nanoswitches might find applications as smart-glue in the field of DNA nanoarchitectures to direct the self-assembly of DNA building blocks (**Figure 45k**), as well as in medicinal chemistry, as smart gates to control the release of drugs from mesoporous silica nanoparticles.^[429] The possibility to implement such a stimuli-responsive unit in an AFM as optomechanical molecular motor to extract net work can also be envisioned.^[430]

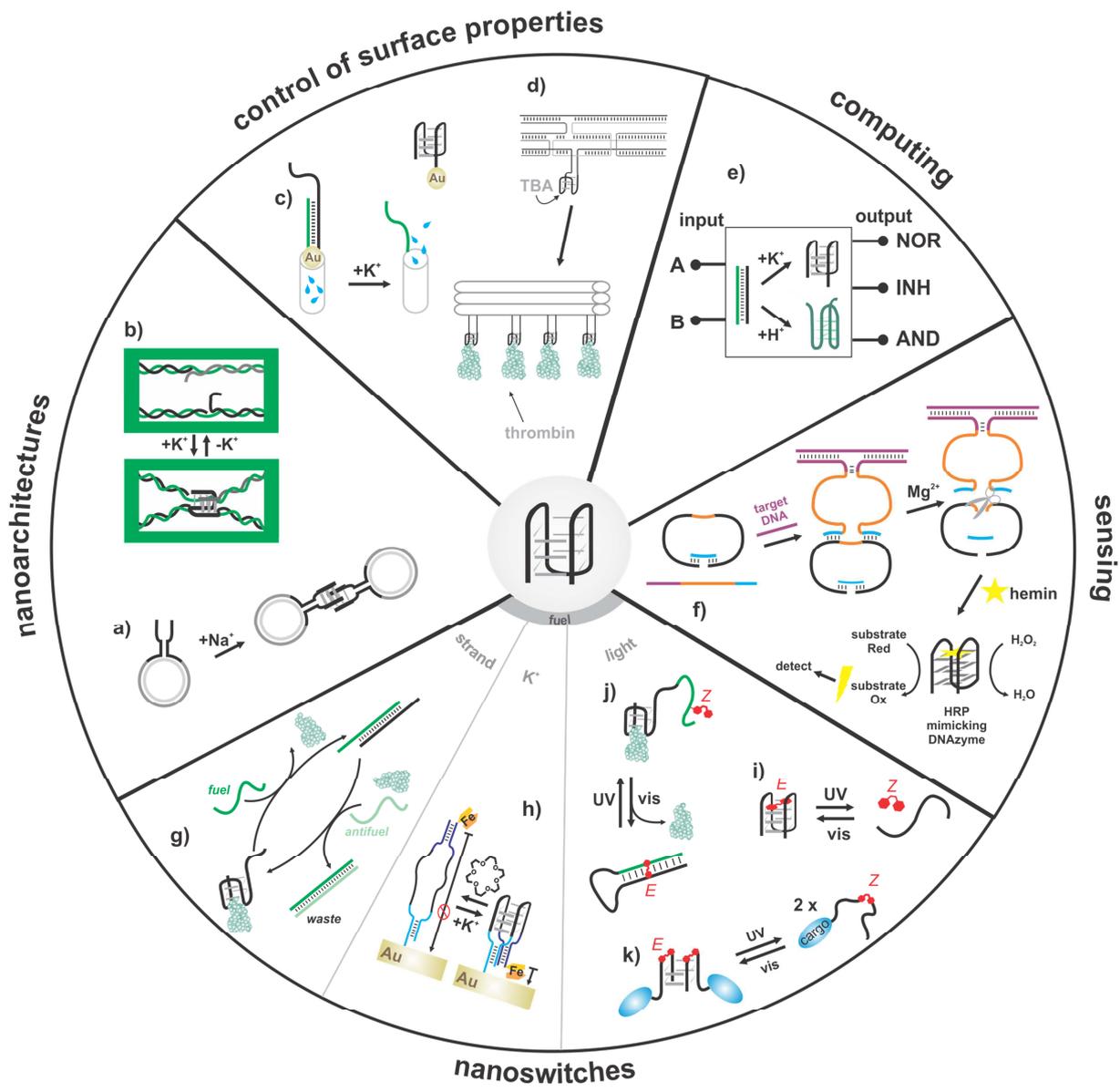


Figure 45 Possible applications of G-quadruplex structures in nanotechnology.

8.1 Research article: Photoresponsive Formation of an Intermolecular Minimal G-Quadruplex Motif

Julie Thevarpadam[#], Irene Bessi[#], Oliver Binas[#], Diana P. N. Gonçalves, Chavdar Slavov, Hendrik R. A. Jonker, Christian Richter, Josef Wachtveitl, Harald Schwalbe, Alexander Heckel.

In press.

[#] These authors contributed equally.

This article describes the structural and photoswitching properties of a novel intermolecular G-quadruplex containing azobenzene units, with potential applications in the field of nanotechnology.

Various bifunctional azobenzene linkers (Az1, Az2 and Az3) were flanked by 1-3 guanine nucleotides at either side and their ability to form intermolecular G-quadruplex structures was investigated. The GG-Az1-GG derivative revealed to be the most promising candidate, able to form a well-defined, structurally homogeneous G-quadruplex and to disrupt it under photocontrol. The structure of the GG-Az1-GG G-quadruplex was determined by NMR and its photoswitching behaviour was characterized by CD, NMR and ultrafast UV-pump/mid-IR-probe spectroscopy.

All the Az derivatives were synthesized and characterized by CD in the group of Prof. Heckel (Goethe University, Frankfurt). The NMR measurements, the NMR data analysis as well as the NMR-based structure calculation were performed in the group of Prof. Schwalbe. The assignment of the GG-Az1-GG derivative was performed by Mr. Binas and the structure calculation by Dr. Jonker. The author of this thesis performed the preliminary screening of the various derivatives and contributed to NMR data analysis, performing the experiments and to all the aspects of writing the manuscript. The ultrafast UV-pump/mid-IR-probe experiments were performed and analyzed in the group of Prof. Wachtveitl (Goethe University, Frankfurt).

Photoresponsive Formation of an Intermolecular Minimal G-Quadruplex Motif

Julie Thevarpadam⁺, Irene Bessi⁺, Oliver Binas⁺, Diana P. N. Gonçalves, Chavdar Slavov, Hendrik R. A. Jonker, Christian Richter, Josef Wachtveitl, Harald Schwalbe,* and Alexander Heckel*

In memory of Gerhard Quinkert

Abstract: The ability of three different bifunctional azobenzene linkers to enable the photoreversible formation of a defined intermolecular two-tetrad G-quadruplex upon UV/Vis irradiation was investigated. Circular dichroism and NMR spectroscopic data showed the formation of G-quadruplexes with K⁺ ions at room temperature in all three cases with the corresponding azobenzene linker in an E conformation. However, only the para-para-substituted azobenzene derivative enables photoswitching between a nonpolymorphic, stacked, tetramolecular G-quadruplex and an unstructured state after E-Z isomerization.

G-Quadruplexes are important DNA secondary structures. The structures form in an intramolecular manner or by association of multiple strands and can exist in many polymorphic forms.^[1–4] G-Quadruplexes are important regulatory elements in the genome^[5,6] but they are also selected in SELEX procedures as versatile aptamers (SELEX = systematic enrichment of ligands by exponential amplification).^[7] For example a simple G-rich DNA 15mer can inhibit blood clotting.^[8] In DNA nanoarchitectures, G-quadruplexes can act as interesting structural scaffolds.^[9–14]

Light is an ideal external trigger signal that can be highly selective and superior to changes in temperature and pH value. Localized irradiation, for example in laser scanning

microscopes, allows for very precise spatiotemporal and dose control,^[15–18] far beyond the precision of the injection of trigger compounds, as recently shown for blood clotting and miRNA.^[19,20]

The ability to control a process by light can be introduced either by using photolabile groups or photoswitches such as azobenzene.^[24,25] The use of photolabile groups has been applied to G-quadruplexes which were either formed or destroyed irreversibly upon light irradiation.^[21–23]

In the G-quadruplex field, Ogasawara and Maeda have used stilbene-type substituents to control the formation of G-quadruplexes by light-induced E-Z isomerization.^[26] Spada et al. controlled the self-assembly of guanosine monomers in a similar fashion.^[27] The formation of G-quadruplexes can also be induced by small molecules that act as molecular chaperons in the absence of cations.^[28–30] Zhou et al. developed azobenzene-containing small-molecule chaperones to regulate G-quadruplex formation.^[31] Tan et al. synthesized an azobenzene-modified antisense DNA linked to a thrombin-binding aptamer to regulate G-quadruplex formation.^[32]

Herein, we present a minimal light-switchable DNA module enabling the formation of an intermolecular and conformationally well-defined G-quadruplex structure with a photoswitchable azobenzene residue as part of the backbone structure (Figure 1).

Azobenzene derivatives **Az1**, **Az2**, and **Az3** (Figure 2a) were employed as photoswitchable linkers between two sets of two consecutive guanosine moieties and were introduced using DNA solid-phase synthesis. The size and substitution patterns on the azo linkers were chosen to offer a suitable balance between the rigidity and flexibility of the overall structure, such that the photoswitch in the E conformation should permit the formation of a G-quadruplex, whereas in the Z conformation no such G-quadruplex formation should be possible. Simple predictions suggested that **Az1** can bridge a distance of 13.2–13.6 Å in the E conformation and 7.7–

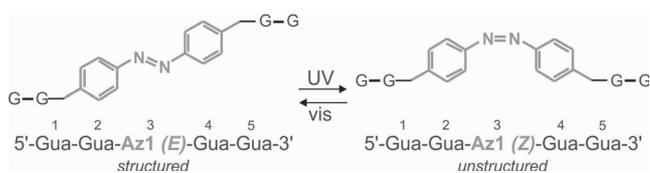


Figure 1. Sequence and numbering of the azobenzene derivative used and characterized in this study, with the azo unit (**Az1**) shown in gray.

[*] J. Thevarpadam,^[†] Dr. D. P. N. Gonçalves, Prof. Dr. A. Heckel
 Goethe University Frankfurt
 Institute for Organic Chemistry and Chemical Biology
 Buchmann Institute for Molecular Life Sciences
 Max-von-Laue-Strasse 9, 60438 Frankfurt (Germany)
 E-mail: heckel@uni-frankfurt.de
 Homepage: <http://photochem.uni-frankfurt.de>

I. Bessi,^[†] O. Binas,^[†] Dr. H. R. A. Jonker, Dr. C. Richter,
 Prof. Dr. H. Schwalbe
 Institute for Organic Chemistry and Chemical Biology
 Center for Biomolecular Magnetic Resonance (BMRZ)
 Max-von-Laue-Strasse 9, 60438 Frankfurt (Germany)
 E-mail: schwalbe@nmr.uni-frankfurt.de
 Homepage: <http://schwalbe.org.chemie.uni-frankfurt.de>

Dr. C. Slavov, Prof. Dr. J. Wachtveitl
 Institute for Physical and Theoretical Chemistry
 Max-von-Laue-Strasse 7, 60438 Frankfurt (Germany)

[†] These authors contributed equally to this work.

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.201510269>.

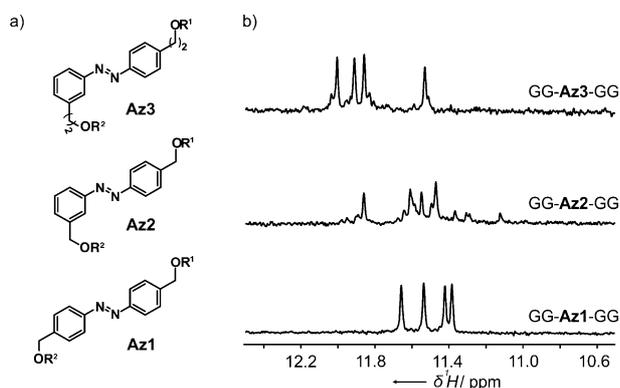


Figure 2. a) The structures of the azo units employed in this study. b) Imino regions of the 1D ^1H NMR spectra of GG-**Az1**-GG, GG-**Az2**-GG, and GG-**Az3**-GG in the presence of KCl (100 mM). Experimental conditions: DNA (50 μM), Tris-HCl buffer (50 mM; pH 7.4), 298 K, 600 MHz.

11.5 Å in the *Z* conformation between the two oxygen atoms adjacent to the azobenzene core. Going from a *para-para* to a *para-meta* substitution pattern in **Az2**, these values change to 11.1–13.6 Å and 9.0–11.9 Å, respectively. The O–O distance at equivalent positions in the narrow grooves in G-quadruplex structures can reach 11–12 Å (see for example PDB 2GKU). We also included the double homologue **Az3** (8.8–14 Å in the *E* conformation and 7.3–12.4 Å in the *Z* conformation). Clearly, minor changes in the structure of the azo linker result in significant changes of the “hinge qualities” of the photoswitch linkers. All of these considerations should be considered with reservation given the highly polymorphic nature of G-quadruplex structures and their respective structural flexibility. We specifically refrained from using longer homologues so as not to dissipate the perturbation induced by the *E-Z* transition into too many internal degrees of freedom.

Initial ^1H NMR (100 mM K^+) characterization of the short azobenzene-linked sequences chosen for this investigation showed signals in the imino region of the spectrum, typical for Hoogsteen-type hydrogen bonds (Figure 2b), suggesting the formation of G-quadruplex structures. Circular dichroism (CD) studies were performed to assess the conformational properties of the modified G-rich sequences. After the addition of 25 mM of K^+ ions, a positive signal around $\lambda = 295$ nm and a negative signal near $\lambda = 260$ nm were detected for all three systems, indicating formation of antiparallel G-quadruplex structures (see Figure 3a for GG-**Az1**-GG and the Supporting Information for GG-**Az2**-GG and GG-**Az3**-GG).^[33,34] Interestingly, ^1H NMR analysis of the different derivatives revealed structural polymorphism for GG-**Az2**-GG and GG-**Az3**-GG which was not detectable by CD. In all three cases, no formation of higher order aggregates was detected (see Figure S58 in the Supporting Information). For details on the thermal stability of the structures formed by the three sequences, please see Figure S5.

Only four ^1H NMR signals for imino groups, indicating a highly symmetric G-quadruplex structure, were detected after addition of K^+ to GG-**Az1**-GG. The ^1H NMR spectra of GG-**Az1**-GG also showed that at a K^+ ion concentration of 25 mM, the intensity of the set of imino signals was already

65% of its maximum value, whereas at a 5 mM K^+ ion concentration, the signal intensity was at 10% of its maximum (Figure 3b). In absence of K^+ ions there is no detectable interaction between the nucleobases, indicating that neither G-quadruplexes nor any other aggregates of the oligonucleotide strands are formed. Additionally, at Na^+ concentrations of up to 500 mM, we did not detect the formation of any secondary structure (Figures 3c and 3d). In this case, G-quadruplex formation was induced only after adding an additional 100 mM of K^+ ions (Figure 3d; Figure S39).

For another recent example of ion selectivity in G-quadruplexes and possible application in nanotechnology, see also Ref. [36]. Cation selectivity has also been reported for the thrombin-binding aptamer G-quadruplex,^[37] whose G-tetrad core adopts the same folding topology as we have determined for GG-**Az1**-GG (see Figure 4 and the corresponding discussion). We propose that the smaller size of the Na^+ ion is not optimal to coordinate all eight O6 atoms in the two-tetrad cavity, which is necessary to keep the GG-**Az1**-GG in a stable quadruplex structure. Furthermore, the azobenzene linkers may introduce additional strain to the ionic channel, leading to the observed ion selectivity.

We investigated structural changes upon photoswitching of the azobenzene units. Two distinguishable sets of CD spectra could be obtained upon irradiation with either UV or visible light (Figure 3e). After an initial irradiation with UV light for 5 min, the CD signal at $\lambda = 295$ nm disappeared completely. Over the course of 30 min, the system reached thermal equilibrium (red data points in Figures 3e–g). Whereas the thermal *Z-E* isomerization occurs over the course of several days (Figure S42), irradiation with visible light for a short period of time (2 min) led to the almost complete recovery of the initial CD signal after 15 min of equilibration (Figure 3f) which could also be paralleled with corresponding NMR experiments (Figure 3h; Figure S60a). ^1H NMR spectroscopy showed the complete disappearance of signals for the imino protons upon irradiation with UV light and their almost complete recovery after irradiation with visible light (Figure 3i, 100 mM of K^+). We found that the degree of recovery after irradiation with visible light is a function of the DNA concentration: although at 50 μM the recovery of signals attributable to a G-quadruplex structure after irradiation with visible light is basically complete, a sample containing 125 μM DNA shows 75% recovery of resonance signals for the G-quadruplex 15 min after irradiation. However, the recovery of the G-quadruplex structure is complete after thermal equilibration (Figure S60a).

We speculate that at DNA concentrations greater than 50 μM , after UV illumination unspecific aggregates are formed (indicated by a broad signal in the aromatic region of the ^1H NMR spectrum) that slowly convert into the quadruplex folded state and/or to the completely unfolded state.

No effects of UV degradation (such as the photo-oxidation of guanine) were detected in the aromatic region of the 1D ^1H NMR spectra after repetitive UV/visible-light irradiation cycles (Figure S60b). Additionally, UV/Vis difference spectra (Figure S6) showed the differential absorption signature typical for G-quadruplexes.^[38]

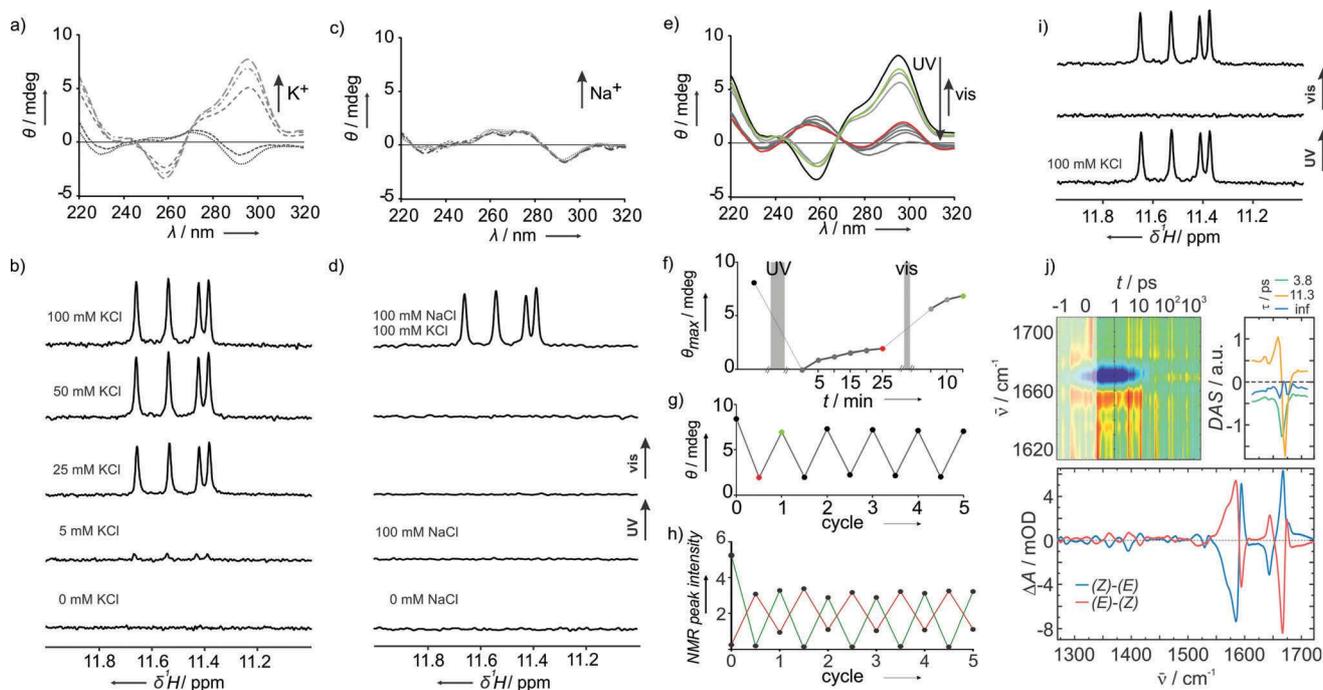


Figure 3. Spectroscopic characterization of the GG-Az1-GG sequence. a) CD spectra recorded with increasing K^+ concentrations (0, 5, 25, 50, 100 mM). b) 1D 1H NMR spectra showing the imino region recorded in the presence of different KCl concentrations. c) CD spectra with increasing concentrations (0, 50, 100, 200, 300, 400, 500 mM) of Na^+ . d) Imino region of the 1D 1H NMR spectra in the presence of increasing amounts of NaCl and KCl. e) CD spectra of GG-Az1-GG with photoirradiation (5 min UV, 2 min visible light) and f) corresponding time course of the signal at $\lambda = 295$ nm under the indicated treatment. Cycling of the photoresponsive structural conversion of GG-Az1-GG by alternate irradiation with UV (15 min) and visible (4 min) light as monitored g) by CD at $\lambda = 295$ nm and h) by 1H NMR ([DNA] = 125 μ M, [KCl] = 100 mM) at $\delta = 11.55$ ppm (green line; monitoring the intensity of the signal for the imino proton of residue G5 in the *E* conformation) and at 6.64 ppm (red line; monitoring the intensity of the signal for the aromatic proton from the azobenzene moiety in the *Z* conformation). Absolute NMR peak intensities are referenced to an internal standard. i) Imino region of the 1D 1H NMR spectra ([DNA] = 50 μ M, [KCl] = 100 mM) showing spectral changes with photoirradiation (30 min UV, 2 min visible light). j) IR difference spectra (bottom); IR transient absorption data recorded from GG-Az1-GG after excitation of the azobenzene moiety at $\lambda = 335$ nm (top left); decay-associated spectra from the global lifetime analysis^[35] of the transient absorption data (top right).

Photoswitching of GG-Az1-GG was also evident in the FTIR difference spectra (Figure 3j, bottom; Figure S53). The azobenzene isomerization and the subsequent disruption of the G-quadruplex structure led to a bathochromic shift ($1675\text{ cm}^{-1} \rightarrow 1667\text{ cm}^{-1}$) of the C=O stretching band^[39–41] and to batho- and hypsochromic shifts of the bands associated with the C=C and C=N purine ring vibrations ($1550\text{--}1590\text{ cm}^{-1}$ range),^[39,41] features reported in the literature to be characteristic of G-quadruplex melting.^[39,41] Ultrafast UV-pump/mid-IR-probe experiments^[42] in the carbonyl stretching vibration range (Figure 3j, top left; Figure S54–55) were performed to investigate the dynamics of the G-quadruplex after *E* azobenzene excitation ($\lambda = 335$ nm). In the first picoseconds after laser excitation, the transient absorption data are dominated by the cooling dynamics of the C=O stretching band ($1645\text{--}1680\text{ cm}^{-1}$) with a lifetime of circa 11 ps. The absorption and bleach bands for the product state become visible after circa 10 ps (see the infinity spectrum in Figure 3j, top right). The last spectrum in the transient absorption data is essentially a *Z–E* IR spectrum of GG-Az1-GG at about 1800 ns. Evidently, this spectrum does not fully match the corresponding FTIR difference spectrum (Figure 3j, bottom), which indicates that despite the nearly instantaneous disruption of the FTIR features of a G-quad-

ruplex, residual conformational dynamics occur on longer time scales.

For studies of GG-Az1-GG derivatives elongated or shortened at the 3'- and 5'-end and for the results obtained with the sequence GGG-Az1-GGG, please see Figures S40/S41 and S57, respectively.

To elucidate the molecular structure of the homogeneously folded GG-Az1-GG, 2D NMR experiments (2D $^1H\text{--}^1H$ NOESY, $^1H\text{--}^{13}C$ HSQC, and $^1H\text{--}^{13}C$ HMBC) were conducted to obtain the complete proton chemical shift assignment. Distance restraints were derived from NOE data and additional angular restraints were obtained from high-resolution $^1H\text{--}^{13}C$ HSQC, 2D $^1H\text{--}^1H$ P.E.COSY, and $^1H\text{--}^{31}P$ TOCSY experiments.^[43] Assignment and *J*-coupling analysis are reported in the Supporting Information.

The folding topology of GG-Az1-GG was determined on the basis of NOESY data. As indicated by intra-tetrad H1–H8 connectivities (Figures S44), GG-Az1-GG adopts a symmetric, antiparallel G-quadruplex structure with tetrads composed as shown in Figure 4 a, b. The edgewise loops containing the azobenzene moieties are located above the G2–G4 tetrad. The anomeric-aromatic region of the NOESY spectrum (Figure S45) indicated a *syn* conformation for the glycosidic bond of residues G1 and G4 and an *anti* conformation for residues G2 and G5.

Intriguingly, the structure of GG-Az1-GG resembles very much the ones of the G-quadruplex formed by the thrombin-binding aptamer (PDB: 148D, NMR; and PDB: 4DII, X-ray)^[37,44] and in the promoter region of the B-raf gene (PDB: 4H29; see Figure S59 for an overlay).^[45]

Hydrogen–deuterium exchange experiments (Figure S46) revealed that imino protons belonging to residues G1 and G5 are protected from solvent exchange. This result suggests a tetrameric G-quadruplex consisting of two dimeric units with the previously defined topology, in which the G1–G5 tetrads of both dimers face each other (Figure 4c). The stacking of two dimeric G-quadruplex units is supported by DOSY data (Figure S47). The structure of the tetramer (Figure 4d) was calculated using ARIA (details in the Supporting Information). The tetrameric arrangement was further supported by the fact that structure calculations as either monomer or dimer led to NOE violations. A significant number of NOEs are unambiguously assigned for the dimer and tetramer (Table 1). The prevalent conformation of the sugar moieties of G1, G2, and G4 is C2'-endo (confirmed by typical strong NOE cross-peaks and ³J coupling constants). Only the 3'-terminal G5 sugar moiety is less well-defined, in part caused by resonance overlap, but manual inspection of the NOE cross-peaks (which are weaker than for the others) also suggest that this sugar possibly interconverts between the C3'-endo and C2'-endo conformations. Therefore no additional torsion angle restraints were included for the G5 nucleoside. The final bundle of structures was refined in explicit water, including three coordinated potassium ions within the tetrads. The structure is well-defined with an average root-mean-square deviation (RMSD) to mean for all atoms of 0.60 Å.

In summary, we have developed a photoswitchable G-quadruplex module and have characterized its photochemical behavior and its 3D conformation. Out of three investigated

Table 1: Statistics of the structure calculation.^[a]

NOE distance restraints	193
Unambiguous NOEs:	146
intra-residue	87
sequential	42
long-range	1
dimer	4
tetramer	12
Ambiguous NOEs:	47
intra-residue	5
intra-monomer	22
monomer or dimer	7
monomer or tetramer	3
dimer or tetramer	8
mono-, di-, or tetramer	2
Distance restraints	14
intra-monomer hydrogen bonds	4
inter-monomer hydrogen bonds	4
potassium site coordination	6
Base planarity	4
intra-monomer	2
inter-monomer	2
Torsion angles	35
backbone	13
β (from ³ J(H5' _{1,2} ,P))	3
glycosidic (χ)	4
sugar pucker	15
Violations	0
distances (>0.3 Å)	0
dihedral angles (>5°)	0
RMSD (average to mean)	
monomer (all atom)	0.55 Å
tetramer (all atom)	0.60 Å

[a] Statistics per monomer.

photoswitchable linkers in a number of sequence contexts, only GG-Az1-GG showed a defined and robust structural behavior. The system shows excellent photocontrol by UV/

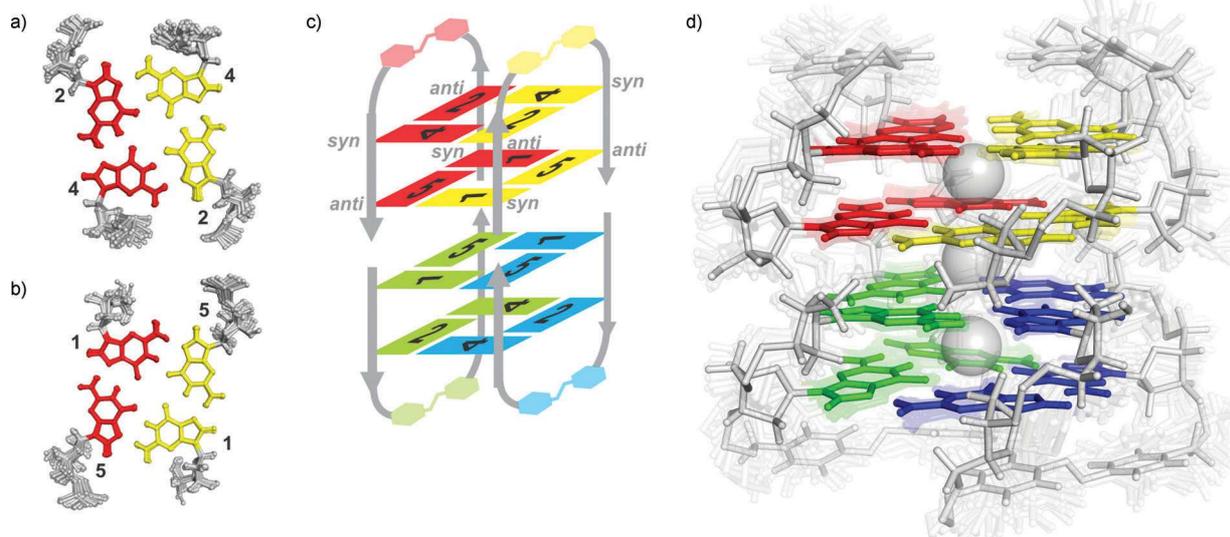


Figure 4. a) The G2–G4 tetrad and b) the G1–G5 tetrad with color coding corresponding both to the structure shown in (c) and to residue numbering given in Figure 1. c) Representation of the structure of the tetramer, with each monomer displayed in a different color. Conformations of the bases (*syn* or *anti*) are indicated. d) NMR solution structure of the G-quadruplex, showing the best representative of the bundle (indicated faintly in the background). K⁺ ions are shown as gray spheres. PDB-code: 2N9Q; BMRB-code: 25915.

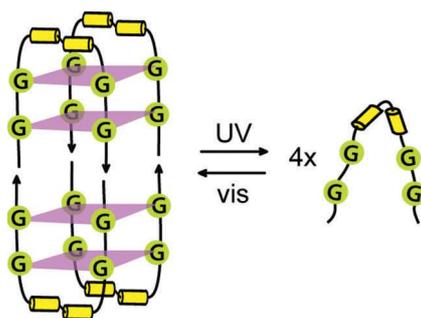
Communications



G-Quadruplexes

J. Thevarpadam, I. Bessi, O. Binas,
D. P. N. Gonçalves, C. Slavov,
H. R. A. Jonker, C. Richter, J. Wachtveitl,
H. Schwalbe,* A. Heckel* — ■■■-■■■

Photoresponsive Formation of an
Intermolecular Minimal G-Quadruplex
Motif



Switch and self-assemble: Three different azobenzene-containing bifunctional guanosine sequences were employed for the construction, upon UV/Vis irradiation, of a defined intermolecular two-tetrad G-quadruplex. The use of one azobenzene derivative enables photoswitching between a stacked, tetramolecular G-quadruplex and an unstructured state after *E*-*Z* isomerization.

German summary

Diese Doktorarbeit beschäftigt sich mit der NMR-Charakterisierung der Struktur und der Faltungsdynamik von DNA-G-Quadruplexen als mögliches therapeutisches Ziel in der Krebstherapie und Baustein für die DNA-basierte Nanotechnologie.

In Gegenwart stabilisierender Kationen wie Na^+ oder K^+ , können G-reiche Sequenzen G-Quadruplexstrukturen durch Stapeln mehrerer G-Tetraden ausbilden. Dabei besteht jede G-Tetrade aus der zyklischen Anordnung von vier Guaninbasen, die durch Hoogsteen-Wasserstoffbrücken stabilisiert wird. Die kanonische B-DNA wurde für lange Zeit als einzige physiologisch relevante DNA-Konformation angesehen; allerdings wurde kürzlich die Rolle der G-Quadruplexstrukturen als regulatorisches Element in vielen biologischen Prozessen erkannt.

Telomer-DNA wird durch die Wiederholung eines G-reichen Motivs gebildet und ihr einzelsträngiger 3'-Überhang ist das Substrat der Telomerase, die in 85-90% aller menschlichen Krebsarten hochreguliert ist. Die Bildung von G-Quadruplexen am 3'-Ende der menschlichen Telomer-DNA wurde nachgewiesen. Die Bildung von G-Quadruplexstrukturen kann die Rekrutierung von Telomerase und Telomer-assoziierten Proteinen, die für das „Capping“ und die Integrität der Chromosomenenden erforderlich sind, beeinflussen. Daher richteten sich die Forschungsbemühungen in letzter Zeit auf kleine Moleküle, die aufgrund ihrer Fähigkeit sich an G-Quadruplexstrukturen zu binden und diese zu stabilisieren, potentielle Krebsmedikamente darstellen.

Jedoch ist aufgrund des hohen Polymorphismus der menschlichen Telomer-DNA die bisher bekannte Strukturinformation für G-Quadruplex-Ligand-Komplexe nur spärlich. Um die Bindungsmechanismen zwischen der G-Quadruplexstruktur und den wechselwirkenden Liganden besser zu verstehen und zu interpretieren, sind weitere Informationen über den Aufbau und die Dynamik der menschlichen Telomer-DNA erforderlich. Außerdem ist ein tieferes Verständnis der Faltungsmechanismen telomerer G-Quadruplexe notwendig, um die Wechselwirkung des 3'-Überhangs der Telomere mit Liganden und Telomer-bindenden Proteinen aufzuklären.

Bioinformatische Vorhersagen ergaben auch in anderen Regionen des Genoms die potentielle Bildung von G-Quadruplexstrukturen, was auf deren mögliche Rolle in der Transkriptionsregulation hindeutet. Insbesondere sind potentielle G-Quadruplex-bildende Sequenzen in Promotorregionen mehrerer Onkogene, im Vergleich zu Tumorsuppressorgenen und Haushaltsgenen, angereichert. Die Effekte der G-Quadruplexbildung und deren Stabilisierung durch kleine Moleküle sind am besten für die Transkription des c-MYC-Onkogens untersucht, dessen Expression in einer Reihe von malignen Erkrankungen dereguliert ist. Es wurde gezeigt, dass die Bindung von kleinen Molekülen an die G-Quadruplexstrukturen, die in der G-reichen Sequenz oberhalb der c-MYC-Promotorregion gebildet werden, eine Herabregulation der c-MYC-Expression zur Folge hat. Die

Entwicklung von kleinen Liganden, die in der Lage sind, die c-MYC-Expression zu unterdrücken, wird als attraktive Strategie für die Krebstherapie angesehen.

Abgesehen von dieser wichtigen Rolle in der Genom-Erhaltung und der Genregulation biologischer Prozesse werden G-Quadruplexstrukturen auch in der Nanotechnologie verwendet, um die Selbstanordnung von Nanoarchitekturen zu lenken und um Nanomaschinen herzustellen.

Der erste Teil der Arbeit (Kapitel 1-5) führt den Leser in die Welt der G-Quadruplexe ein.

Die wichtigsten Merkmale der klassischen Watson-Crick-Doppelhelix und alternativer nicht-B-DNA-Strukturen sind in Kapitel 1 dargestellt. Viele verschiedene Basenpaarungssysteme sind neben dem kanonischen Watson-Crick-Motiv möglich, wodurch die strukturelle Komplexität der DNA erweitert wird. Nicht kanonische Basenpaarungen, wie die Hoogsteen-Wasserstoffbrückenbindung, ermöglichen die Ausbildung von Triplets und Quartetts, die die Bausteine der Triplex- bzw. Quadruplex-Strukturen sind.

Die strukturellen Eigenschaften von DNA G-Quadruplexen sind ausführlich in Kapitel 2 dargestellt.

G-Quadruplexstrukturen sind extrem polymorph, in Bezug auf die Strangorientierung, Schlaufengeometrie, Breite der Furchen und Anordnung der glycosidischen Torsionswinkel. Die verschiedenen Strukturelemente als auch die Koordinationsgeometrien verschiedener Kationen werden hier vorgestellt. Ein besonderer Schwerpunkt liegt auf der Vielfalt der Konformationen, die für die Telomer-DNA-G-Quadruplexe beschrieben worden sind.

Kapitel 3 beschreibt die biologische Rolle von G-Quadruplexstrukturen im Genom. Nach einer Einführung in die Architektur der Telomer-DNA und der mit dieser interagierenden Proteine wird der Mechanismus der Telomerverlängerung, der durch die Telomerase katalysiert wird, und seine Auswirkungen auf Krebs diskutiert. Das Auftreten von G-Quadruplexstrukturen in funktionellen Regionen des Genoms, wie zum Beispiel in Promotorregionen von Onkogenen, und ihre mögliche Rolle bei der Regulation der Gentranskription werden dann in dem zweiten Teil des Kapitels beschrieben.

Das Potential des G-Quadruplexes als neuartiger Rezeptor für Krebsmedikamente wird in Kapitel 4 untersucht und die vorgeschlagenen Anti-Krebs-Mechanismen Wirkungsweisen für G-Quadruplexstrukturen-stabilisierende Liganden werden diskutiert. RNA-G-Quadruplexe und ihre mutmaßliche Rolle bei der Genregulation auf der Ebene der Translation werden am Ende des Kapitels kurz dargestellt.

Eine allgemeine Übersicht über die NMR-Methoden, mit denen G-Quadruplexstrukturen untersucht werden können, wird in Kapitel 5 präsentiert. Ebenso ist hier der Versuchsaufbau für die Echtzeit-NMR-Untersuchungen der G-Quadruplex-Faltung beschrieben.

Der zweite Teil der Arbeit (Kapitel 6-8), der kumulative Teil, umfasst die Originalpublikationen in drei Kapiteln, thematisch gegliedert:

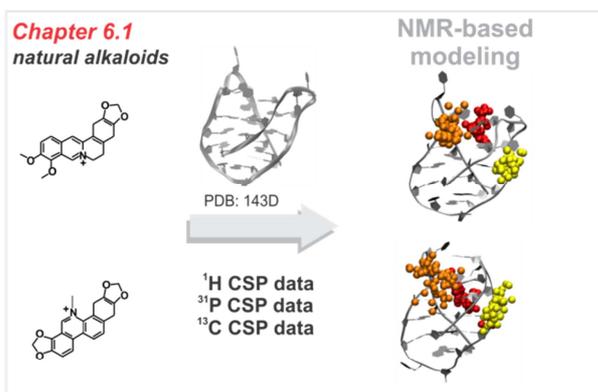
- mit G-Quadruplexen wechselwirkende Liganden (Kapitel 6),
- Faltungskinetiken von G-Quadruplexen (Kapitel 7),

- G-Quadruplexe in der Nanotechnologie (Kapitel 8).

Eine kurze Einführung in das Thema zeigt den Rahmen für die darauf folgenden Publikationen in jedem Kapitel auf. Der persönliche Beitrag des Autors dieser Arbeit zu jeder Publikation ist in der Einleitung zum jeweiligen Artikel angegeben.

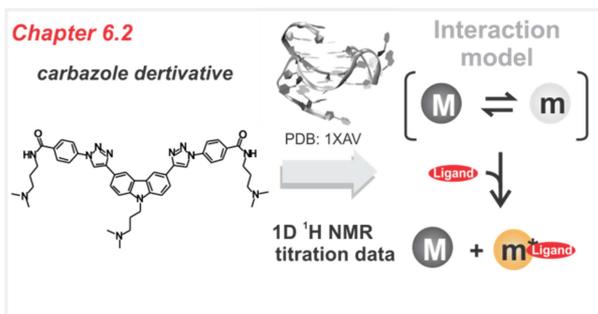
Der aktuelle Stand der Forschung für an G-Quadruplexstrukturen gebundene kleine Moleküle wird zu Beginn von Kapitel 6 beschreiben, einschließlich einer Zusammenfassung der bisher zur Verfügung stehenden Strukturen von G-Quadruplex-Ligand-Komplexen.

Die in den Kapiteln 6.1-6.3 vorgestellten Publikationen befassen sich mit der Aufklärung von Interaktionsformen der DNA-G-Quadruplexe und ausgewählten Liganden mit potentiellen therapeutischen Anwendungen. Das Bindungsverhalten von zwei natürlichen Alkaloiden (Berberin und Sanguinarin) an Telomer-G-Quadruplexe wird in Kapitel 6.1 untersucht. Die Fähigkeit von Carbazol- bzw. Diguanosin-Derivaten (in der Gruppe von Prof. Dash, IISER, Kolkata synthetisiert) mit dem c-MYC-G-Quadruplex zu interagieren und die c-MYC-Expression herabzuregulieren, wird in Kapitel 6.2 bzw. Kapitel 6.3 untersucht.



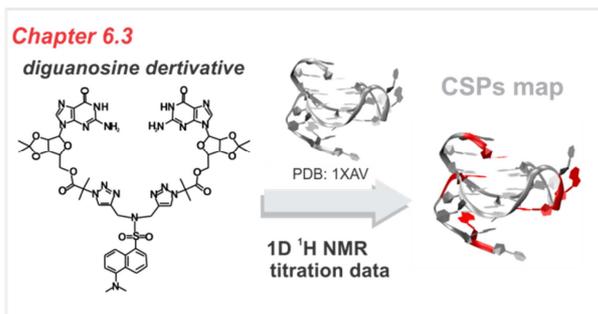
Ergebnisse:

- Identifikation einer hoch-stöchiometrischen Bindung mit Selbstassoziation des Liganden induziert durch die DNA-Bindung
- Abbildung der Bindungsstellen mittels NMR-gesteuertem Docking



Ergebnisse:

- Vorschlag eines Interaktionsmodells (konformationelle Auswahl) auf der Grundlage von 1D-¹H-NMR-Titrationsdaten



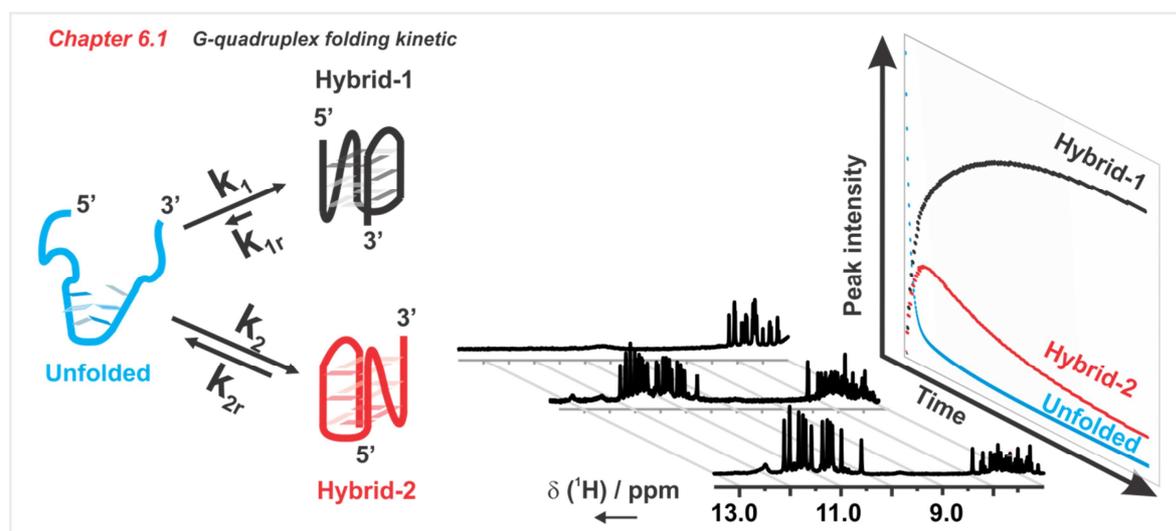
Ergebnisse:

- Abbildung der Interaktionsschnittstelle zwischen dem Liganden und dem c-MYC-G-Quadruplex mittels ¹H-chemical-shift-perturbation (CSP).

Die Energielandschaft der menschlichen Telomer-G-Quadruplex-Strukturen in Bezug auf die bisher publizierten Ergebnisse kinetischer Experimente und Moleküldynamiksimulationen wird in Kapitel 7 diskutiert. Bis heute gibt es keinen allgemeinen Konsens über den Faltungsweg eines unimolekularen, menschlichen Telomer-G-Quadruplexes, insbesondere aufgrund des Fehlens von Daten über die an der Faltung beteiligten Spezies mit atomarer Auflösung. Kapitel 7.1 präsentiert die erste Echtzeit-NMR-Untersuchung der Faltungskinetiken eines menschlichen Telomer-G-Quadruplexes.

Ergebnisse:

- Anwendung von Echtzeit-NMR um einen Einblick in die Faltungswege eines menschlichen Telomer-G-Quadruplexes unter atomarer Auflösung zu erhalten
- Vorschlag eines Faltungsmechanismus (kinetische Partitionierung) unter Einbeziehung langlebiger Zwischenzustände
- Charakterisierung der Faltungstopologie von den an der Faltung beteiligten Spezies

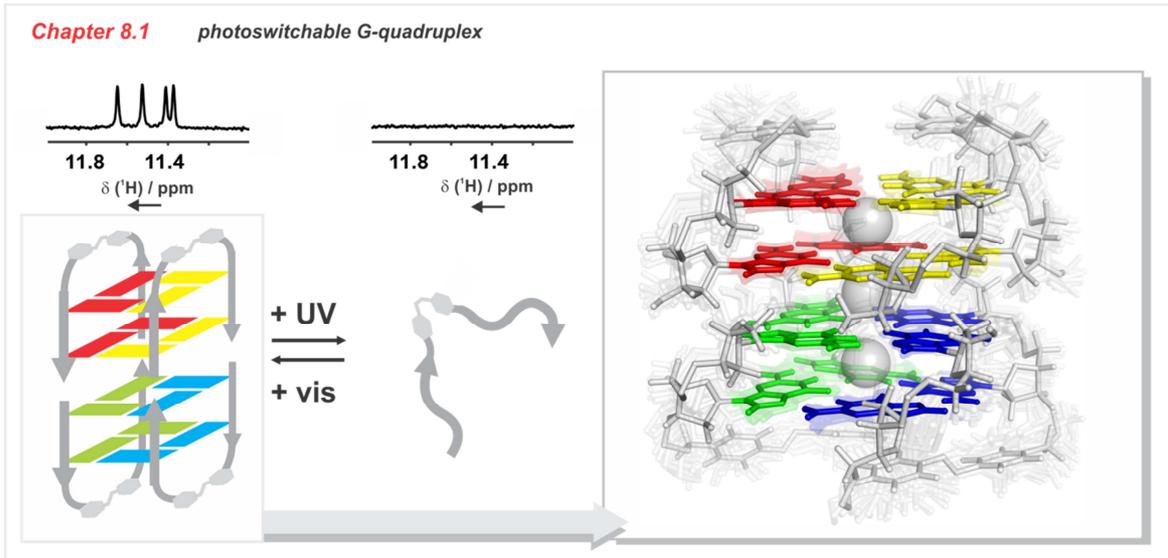


Das letzte Kapitel dieser Arbeit (Kapitel 8) beschreibt das Potenzial der G-Quadruplexstrukturen als Bausteine in der Nanotechnologie. Nach einer kurzen Darstellung der zusätzlichen Möglichkeiten, die alternative nicht-B-DNA-Strukturen zur Programmierung von Nanomaterialien haben, wird eine Reihe von Anwendungen der G-Quadruplexstrukturen in verschiedenen Bereichen der Nanotechnologie beschrieben.

Der in Kapitel 8.1 vorgestellte Artikel untersucht die strukturellen und fotoschaltbaren Eigenschaften eines neuartigen intermolekularen Azobenzol-enthaltenden G-Quadruplexes, der in der Gruppe von Prof. Heckel (Goethe-Universität Frankfurt) synthetisiert wurde.

Ergebnisse:

- Identifizierung eines Azobenzolderivats, welches eine homogene intermolekulare G-Quadruplexstruktur bildet,
- Charakterisierung seines fotoschaltbaren Verhaltens
- Bestimmung der NMR-Struktur des G-Quadruplexes



Bibliography

- [1] A. Ghosh, M. Bansal, *Acta Crystallogr. D. Biol. Crystallogr.* **2003**, *59*, 620–6.
- [2] R. Wing, H. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura, R. E. Dickerson, *Nature* **1980**, *287*, 755–8.
- [3] A. H. Wang, G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. van der Marel, A. Rich, *Nature* **1979**, *282*, 680–6.
- [4] S. M. Mirkin, *Front. Biosci.* **2008**, *13*, 1064–71.
- [5] P. S. Ho, M. Carter, in *DNA Replication-Current Adv.* (Ed.: Herve Seligmann), **2011**.
- [6] L. A. Yatsunyk, O. Mendoza, J.-L. Mergny, *Acc. Chem. Res.* **2014**, *47*, 1836–44.
- [7] C. Altona, M. Sundaralingam, *J. Am. Chem. Soc.* **1972**, *94*, 8205–12.
- [8] S. Nozinovic, C. Richter, J. Rinnenthal, B. Fürtig, E. Duchardt-Ferner, J. E. Weigand, H. Schwalbe, *J. Am. Chem. Soc.* **2010**, *132*, 10318–29.
- [9] J. D. WATSON, F. H. CRICK, *Nature* **1953**, *171*, 737–8.
- [10] J. R. Kiefer, C. Mao, J. C. Braman, L. S. Beese, *Nature* **1998**, *391*, 304–7.
- [11] A. Rich, S. Zhang, *Nat. Rev. Genet.* **2003**, *4*, 566–572.
- [12] G. P. Schroth, P. J. Chou, P. S. Ho, *J. Biol. Chem.* **1992**, *267*, 11846–55.
- [13] R. R. Sinden, T. J. Kochel, *Biochemistry* **1987**, *26*, 1343–50.
- [14] R. E. Dickerson, M. Bansal, C. R. Calladine, S. Diekmann, W. N. Hunter, O. Kennard, E. von Kitzing, R. Lavery, H. C. Nelson, W. K. Olson, et al., *J. Mol. Biol.* **1989**, *205*, 787–791.
- [15] M. A. el Hassan, C. R. Calladine, *J. Mol. Biol.* **1995**, *251*, 648–64.
- [16] X.-J. Lu, W. K. Olson, *Nucleic Acids Res.* **2003**, *31*, 5108–21.
- [17] X.-J. Lu, W. K. Olson, *Nat. Protoc.* **2008**, *3*, 1213–27.
- [18] W. K. Olson, M. Bansal, S. K. Burley, R. E. Dickerson, M. Gerstein, S. C. Harvey, U. Heinemann, X. J. Lu, S. Neidle, Z. Shakked, et al., *J. Mol. Biol.* **2001**, *313*, 229–37.
- [19] K. Hoogsteen, *Acta Crystallogr.* **1959**, *12*, 822–823.
- [20] E. N. Nikolova, E. Kim, A. A. Wise, P. J. O'Brien, I. Andricioaei, H. M. Al-Hashimi, *Nature* **2011**, *470*, 498–502.
- [21] M. Kitayner, H. Rozenberg, R. Rohs, O. Suad, D. Rabinovich, B. Honig, Z. Shakked, *Nat. Struct. Mol. Biol.* **2010**, *17*, 423–9.
- [22] R. Cox, S. M. Mirkin, *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 5237–42.
- [23] S. Rhee, Z. j Han, K. Liu, H. T. Miles, D. R. Davies, *Biochemistry* **1999**, *38*, 16810–5.
- [24] M. Cooney, G. Czernuszewicz, E. H. Postel, S. J. Flint, M. E. Hogan, *Science* **1988**, *241*, 456–9.
- [25] A. Mukherjee, K. M. Vasquez, *Biochimie* **2011**, *93*, 1197–208.
- [26] L. Cerofolini, J. Amato, A. Giachetti, V. Limongelli, E. Novellino, M. Parrinello, M. Fragai, A. Randazzo, C. Luchinat, *Nucleic Acids Res.* **2014**, *42*, 13393–404.
- [27] M. Gellert, M. Lipsett, D. Davies, *Proc. Natl. Acad. Sci. U. S. A.* **1962**, *48*, 2013–2018.
- [28] K. Gehring, J. L. Leroy, M. Gueron, *Nature* **1993**, *363*, 561–565.
- [29] A. L. Lieblein, M. Krämer, A. Dreuw, B. Fürtig, H. Schwalbe, *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 4067–70.
- [30] A. T. Phan, J. L. Leroy, *J. Biomol. Struct. Dyn.* **2000**, *17 Suppl 1*, 245–51.
- [31] A. T. Phan, M. Guéron, J. L. Leroy, *J. Mol. Biol.* **2000**, *299*, 123–44.
- [32] K. Guo, V. Gokhale, L. H. Hurley, D. Sun, *Nucleic Acids Res.* **2008**, *36*, 4598–608.
- [33] H.-J. Kang, S. Kendrick, S. M. Hecht, L. H. Hurley, *J. Am. Chem. Soc.* **2014**, *136*, 4172–85.
- [34] D. Sun, L. H. Hurley, *J. Med. Chem.* **2009**, *52*, 2863–74.
- [35] S. Kendrick, L. H. Hurley, *Pure Appl. Chem.* **2010**, *82*, 1609–1621.

- [36] Y. Dong, Z. Yang, D. Liu, *Acc. Chem. Res.* **2014**, *47*, 1853–60.
- [37] I. Bang, *Biochem. Z.* **1910**, *26*, 293–311.
- [38] T. J. Pinnavaia, C. L. Marshall, C. M. Mettler, C. L. Fisk, H. T. Miles, E. D. Becker, *J. Am. Chem. Soc.* **1978**, *100*, 3625–3627.
- [39] J. T. Davis, G. P. Spada, *Chem. Soc. Rev.* **2007**, *36*, 296–313.
- [40] E. Henderson, C. C. Hardin, S. K. Walk, I. Tinoco, E. H. Blackburn, *Cell* **1987**, *51*, 899–908.
- [41] D. Sen, W. Gilbert, *Nature* **1988**, *334*, 364–6.
- [42] J. R. Williamson, M. K. Raghuraman, T. R. Cech, *Cell* **1989**, *59*, 871–80.
- [43] N. V. Hud, J. Plavec, in *Quadruplex Nucleic Acids* (Eds.: S. Neidle, S. Balasubramanian), RSC, **2006**, pp. 100–130.
- [44] M. Borzo, C. Detellier, P. Laszlo, A. Paris, *J. Am. Chem. Soc.* **1980**, *102*, 1124–1134.
- [45] C. Detellier, P. Laszlo, *J. Am. Chem. Soc.* **1980**, *102*, 1135–1141.
- [46] J. Sponer, J. Leszczynski, P. Hobza, *Biopolymers* **2001**, *61*, 3–31.
- [47] M. Rueda, F. J. Luque, M. Orozco, *J. Am. Chem. Soc.* **2006**, *128*, 3608–19.
- [48] G. A. L. Bare, B. Liu, J. C. Sherman, *J. Am. Chem. Soc.* **2013**, *135*, 11985–9.
- [49] A. Wong, G. Wu, *J. Am. Chem. Soc.* **2003**, *125*, 13895–905.
- [50] M. L. Gill, S. A. Strobel, J. P. Loria, *Nucleic Acids Res.* **2006**, *34*, 4506–14.
- [51] M. L. Gill, S. A. Strobel, J. P. Loria, *J. Am. Chem. Soc.* **2005**, *127*, 16723–32.
- [52] B. I. Kankia, L. A. Marky, *J. Am. Chem. Soc.* **2001**, *123*, 10799–804.
- [53] I. M. Pedroso, L. F. Duarte, G. Yanez, A. M. Baker, T. M. Fletcher, *Biochem. Biophys. Res. Commun.* **2007**, *358*, 298–303.
- [54] A. C. Fyfe, P. W. Dunten, M. M. Martick, W. G. Scott, *J. Mol. Biol.* **2015**, *427*, 2205–19.
- [55] D. Zhang, T. Huang, P. S. Lukeman, P. J. Paukstelis, *Nucleic Acids Res.* **2014**, *42*, 13422–13429.
- [56] F. Kotch, J. Fettingler, J. Davis, *Org. Lett.* **2000**, *2*, 3277–80.
- [57] P. R. Majhi, R. H. Shafer, *Biopolymers* **2006**, *82*, 558–69.
- [58] M. P. . P. G. N. . H. P. . N. S. Lee, *J. Am. Chem. Soc.* **2007**, *129*, 10106–10107.
- [59] I. C. M. Kwan, A. Wong, Y.-M. She, M. E. Smith, G. Wu, *Chem. Commun. (Camb)*. **2008**, 682–4.
- [60] C. C. Hardin, A. G. Perry, K. White, *Biopolymers* **2000**, *56*, 147–194.
- [61] K. Phillips, Z. Dauter, A. I. H. Murchie, D. M. J. Lilley, B. Luisi, *J. Mol. Biol.* **1997**, *273*, 171–182.
- [62] R. D. Shannon, *Acta Crystallogr. Sect. A* **1976**, *32*, 751–767.
- [63] A. N. Lane, J. B. Chaires, R. D. Gray, J. O. Trent, *Nucleic Acids Res.* **2008**, *36*, 5482–5515.
- [64] R. D. Gray, J. B. Chaires, *Biophys. Chem.* **2011**, *159*, 205–9.
- [65] N. V Hud, F. W. Smith, F. A. Anet, J. Feigon, *Biochemistry* **1996**, *35*, 15383–90.
- [66] A. J. Dingley, R. D. Peterson, S. Grzesiek, J. Feigon, *J. Am. Chem. Soc.* **2005**, *127*, 14466–14472.
- [67] P. Sket, M. Crnugelj, W. Kozminski, J. Plavec, *Org. Biomol. Chem.* **2004**, *2*, 1970–3.
- [68] P. Podbevsek, N. V Hud, J. Plavec, *Nucleic Acids Res.* **2007**, *35*, 2554–63.
- [69] P. Podbevsek, P. Sket, J. Plavec, *J. Am. Chem. Soc.* **2008**, *130*, 14287–93.
- [70] M. Trajkovski, P. Sket, J. Plavec, *Org. Biomol. Chem.* **2009**, *7*, 4677–84.
- [71] P. Sket, J. Plavec, *J. Am. Chem. Soc.* **2007**, *129*, 8794–800.
- [72] P. Sket, J. Plavec, *J. Am. Chem. Soc.* **2010**, *132*, 12724–32.
- [73] P. Šket, A. Virgilio, V. Esposito, A. Galeone, J. Plavec, *Nucleic Acids Res.* **2012**, *40*, 11047–57.
- [74] D. Sen, W. Gilbert, *Nature* **1990**, *344*, 410–4.
- [75] M. James L. Lewis, III, “The Merck Manual for Health Care Professionals,” can be found under <http://www.merckmanuals.com/professional/endocrine-and-metabolic-disorders/fluid-metabolism/water-and-sodium-balance>, **2014**.

- [76] J. Dai, T. S. Dexheimer, D. Chen, M. Carver, A. Ambrus, R. A. Jones, D. Yang, *J. Am. Chem. Soc.* **2006**, *128*, 1096–8.
- [77] R. D. Gray, J. Li, J. B. Chaires, *J. Phys. Chem. B* **2009**, *113*, 2676–83.
- [78] R. D. Gray, L. Petraccone, J. O. Trent, J. B. Chaires, *Biochemistry* **2010**, *49*, 179–94.
- [79] P. Sket, M. Crnugelj, J. Plavec, *Nucleic Acids Res.* **2005**, *33*, 3691–7.
- [80] C. Cáceres, G. Wright, C. Gouyette, G. Parkinson, J. A. Subirana, *Nucleic Acids Res.* **2004**, *32*, 1097–102.
- [81] P. Schultze, N. V. Hud, F. W. Smith, J. Feigon, *Nucleic Acids Res.* **1999**, *27*, 3018–28.
- [82] Z.-F. Wang, M.-H. Li, S.-T. D. Hsu, T.-C. Chang, *Nucleic Acids Res.* **2014**, *42*, 4723–33.
- [83] J. Dickerhoff, K. Weisz, *Angew. Chem. Int. Ed. Engl.* **2015**, *54*, 5588–91.
- [84] V. V. Cheong, C. J. Lech, B. Heddi, A. T. Phan, *Angew. Chem. Int. Ed. Engl.* **2015**, DOI 10.1002/anie.201507034.
- [85] C. J. Lech, B. Heddi, A. T. Phan, *Nucleic Acids Res.* **2013**, *41*, 2034–46.
- [86] N. Zhang, A. Gorin, A. Majumdar, A. Kettani, N. Chernichenko, E. Skripkin, D. J. Patel, *J. Mol. Biol.* **2001**, *311*, 1063–79.
- [87] A. Bugaut, S. Balasubramanian, *Biochemistry* **2008**, *47*, 689–97.
- [88] A. Guédin, J. Gros, P. Alberti, J.-L. Mergny, *Nucleic Acids Res.* **2010**, *38*, 7858–68.
- [89] J. T. Nielsen, K. Arar, M. Petersen, *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 3099–103.
- [90] V. Kuryavyi, D. J. Patel, *Structure* **2010**, *18*, 73–82.
- [91] M. Adrian, D. J. Ang, C. J. Lech, B. Heddi, A. Nicolas, A. T. Phan, *J. Am. Chem. Soc.* **2014**, *136*, 6297–305.
- [92] P. Agrawal, C. Lin, R. I. Mathad, M. Carver, D. Yang, *J. Am. Chem. Soc.* **2014**, *136*, 1750–3.
- [93] M. Webba da Silva, M. Trajkovski, Y. Sannohe, N. Ma'ani Hessari, H. Sugiyama, J. Plavec, *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 9167–70.
- [94] A. O. K. K. M. L. H. K. S. U. S. K. M. Matsugami, *J. Mol. Biol.* **2001**, *313*, 255–269.
- [95] R. I. Mathad, E. Hatzakis, J. Dai, D. Yang, *Nucleic Acids Res.* **2011**, *39*, 9023–33.
- [96] A. Guedin, P. Alberti, J.-L. Mergny, *Nucleic Acids Res.* **2009**, *37*, 5559–5567.
- [97] M. Webba da Silva, *Chemistry* **2007**, *13*, 9738–45.
- [98] W. J. Chung, B. Heddi, E. Schmitt, K. W. Lim, Y. Mechulam, A. T. Phan, *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 2729–33.
- [99] A. T. Phan, *Febs J.* **2010**, *277*, 1107–1117.
- [100] J. X. Dai, M. Carver, D. Z. Yang, *Biochimie* **2008**, *90*, 1172–1183.
- [101] D. Renciuik, I. Kejnovska, P. Skolakova, K. Bednarova, J. Motlova, M. Vorlickova, *Nucleic Acids Res.* **2009**, *37*, 6625–6634.
- [102] R. D. Gray, J. Li, J. B. Chaires, *J. Phys. Chem. B* **2009**, *113*, 2676–83.
- [103] A. Marchand, V. Gabelica, *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 1146–1154.
- [104] P. S. Shirude, S. Balasubramanian, *Biochimie* **2008**, *90*, 1197–1206.
- [105] A. C. Krüger, V. Birkedal, *Methods* **2013**, *64*, 36–42.
- [106] V. Singh, M. Azarkh, T. E. Exner, J. S. Hartig, M. Drescher, *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 9728–30.
- [107] I. T. Holder, M. Drescher, J. S. Hartig, *Bioorg. Med. Chem.* **2013**, *21*, 6156–61.
- [108] R. D. Gray, L. Petraccone, R. Buscaglia, J. B. Chaires, *Methods Mol. Biol.* **2010**, *608*, 121–36.
- [109] N. T. Dao, R. Haselsberger, M.-E. Michel-Beyerle, A. T. Phan, *FEBS Lett.* **2011**, *585*, 3969–77.
- [110] M. Bončina, J. Lah, I. Prislán, G. Vesnaver, *J. Am. Chem. Soc.* **2012**, *134*, 9657–63.
- [111] Y. Wang, D. J. Patel, *Structure* **1993**, *1*, 263–282.
- [112] A. Ambrus, D. Chen, J. X. Dai, T. Bialis, R. A. Jones, D. Z. Yang, *Nucleic Acids Res.* **2006**, *34*, 2723–2735.
- [113] G. N. Parkinson, M. P. H. Lee, S. Neidle, *Nature* **2002**, *417*, 876–880.
- [114] B. Heddi, A. T. Phan, *J. Am. Chem. Soc.* **2011**, *133*, 9824–33.

- [115] J. X. Dai, C. Punchihewa, A. Ambrus, D. Chen, R. A. Jones, D. Z. Yang, *Nucleic Acids Res.* **2007**, *35*, 2440–2450.
- [116] J. X. Dai, M. Carver, C. Punchihewa, R. A. Jones, D. Z. Yang, *Nucleic Acids Res.* **2007**, *35*, 4927–4940.
- [117] A. T. Phan, V. Kuryavyi, K. N. Luu, D. J. Patel, *Nucleic Acids Res.* **2007**, *35*, 6517–25.
- [118] K. W. Lim, S. Amrane, S. Bouaziz, W. Xu, Y. Mu, D. J. Patel, K. N. Luu, A. T. Phan, *J. Am. Chem. Soc.* **2009**, *131*, 4301–9.
- [119] Z. Zhang, J. Dai, E. Veliath, R. A. Jones, D. Yang, *Nucleic Acids Res.* **2010**, *38*, 1009–21.
- [120] K. W. Lim, V. C. M. Ng, N. Martín-Pintado, B. Heddi, A. T. Phan, *Nucleic Acids Res.* **2013**, *41*, 10556–62.
- [121] T. Mashimo, H. Sugiyama, *Nucleic Acids Symp. Ser. (Oxf.)* **2007**, 239–40.
- [122] A. Ortega, D. Amorós, J. García de la Torre, *Biophys. J.* **2011**, *101*, 892–8.
- [123] J. Li, J. J. Correia, L. Wang, J. O. Trent, J. B. Chaires, *Nucleic Acids Res.* **2005**, *33*, 4649–59.
- [124] H.-X. Zhou, G. Rivas, A. P. Minton, *Annu. Rev. Biophys.* **2008**, *37*, 375–97.
- [125] M. C. Miller, R. Buscaglia, J. B. Chaires, A. N. Lane, J. O. Trent, *J. Am. Chem. Soc.* **2010**, *132*, 17105–7.
- [126] R. Hansel, F. Lohr, S. Foldynova-Trantirkova, E. Bamberg, L. Trantirek, V. Dotsch, *Nucleic Acids Res.* **2011**, *39*, 5768–5775.
- [127] R. Buscaglia, M. C. Miller, W. L. Dean, R. D. Gray, A. N. Lane, J. O. Trent, J. B. Chaires, *Nucleic Acids Res.* **2013**, *41*, 7934–46.
- [128] D. Miyoshi, T. Fujimoto, N. Sugimoto, *Top. Curr. Chem.* **2013**, *330*, 87–110.
- [129] Y. Xue, Z. Kan, Q. Wang, Y. Yao, J. Liu, Y. Hao, Z. Tan, *J. Am. Chem. Soc.* **2007**, *129*, 11185–91.
- [130] Z.-F. Wang, T.-C. Chang, *Nucleic Acids Res.* **2012**, *40*, 8711–20.
- [131] Y. Xu, T. Ishizuka, K. Kurabayashi, M. Komiyama, *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 7833–6.
- [132] L. Petraccone, C. Spink, J. O. Trent, N. C. Garbett, C. S. Mekmaysy, C. Giancola, J. B. Chaires, *J. Am. Chem. Soc.* **2011**, *133*, 20951–61.
- [133] L. Petraccone, *Top. Curr. Chem.* **2013**, *330*, 23–46.
- [134] J. Abraham Punnoose, Y. Cui, D. Koirala, P. M. Yangyuoru, C. Ghimire, P. Shrestha, H. Mao, *J. Am. Chem. Soc.* **2014**, *136*, 18062–18069.
- [135] L. Petraccone, N. C. Garbett, J. B. Chaires, J. O. Trent, *Biopolymers* **2010**, *93*, 533–548.
- [136] L. Xu, S. Feng, X. Zhou, *Chem. Commun. (Camb.)* **2011**, *47*, 3517–9.
- [137] R. Hänsel, F. Löhr, L. Trantirek, V. Dötsch, *J. Am. Chem. Soc.* **2013**, *135*, 2816–24.
- [138] N. Maizels, L. T. Gray, *PLoS Genet.* **2013**, *9*, e1003468.
- [139] J. L. Huppert, S. Balasubramanian, *Nucleic Acids Res.* **2005**, *33*, 2908–2916.
- [140] A. K. Todd, M. Johnston, S. Neidle, *Nucleic Acids Res.* **2005**, *33*, 2901–7.
- [141] P. Rawal, V. B. R. Kummarasetti, J. Ravindran, N. Kumar, K. Halder, R. Sharma, M. Mukerji, S. K. Das, S. Chowdhury, *Genome Res.* **2006**, *16*, 644–55.
- [142] P. Hazel, J. Huppert, S. Balasubramanian, S. Neidle, *J. Am. Chem. Soc.* **2004**, *126*, 16405–15.
- [143] O. Stegle, L. Payet, J.-L. Mergny, D. J. C. MacKay, J. H. Leon, *Bioinformatics* **2009**, *25*, i374–82.
- [144] H. M. Wong, O. Stegle, S. Rodgers, J. L. Huppert, *J. Nucleic Acids* **2010**, *2010*, DOI 10.4061/2010/564946.
- [145] O. Kikin, L. D'Antonio, P. S. Bagga, *Nucleic Acids Res.* **2006**, *34*, W676–82.
- [146] J. L. Huppert, *Biochimie* **2008**, *90*, 1140–8.
- [147] J. Eddy, N. Maizels, *Nucleic Acids Res.* **2006**, *34*, 3887–96.
- [148] J. L. Huppert, S. Balasubramanian, *Nucleic Acids Res.* **2007**, *35*, 406–13.
- [149] G. E. Crawford, I. E. Holt, J. Whittle, B. D. Webb, D. Tai, S. Davis, E. H. Margulies, Y. Chen, J. A. Bernat, D. Ginsburg, et al., *Genome Res.* **2006**, *16*, 123–31.

Bibliography

- [150] H. M. Wong, J. L. Huppert, *Mol. Biosyst.* **2009**, *5*, 1713–9.
- [151] H. J. Lipps, D. Rhodes, *Trends Cell Biol.* **2009**, *19*, 414–422.
- [152] G. Biffi, D. Tannahill, J. McCafferty, S. Balasubramanian, *Nat. Chem.* **2013**, *5*, 182–6.
- [153] C. Schaffitzel, I. Berger, J. Postberg, J. Hanes, H. J. Lipps, A. Plückthun, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 8572–7.
- [154] R. Rodriguez, K. M. Miller, J. V Forment, C. R. Bradshaw, M. Nikan, S. Britton, T. Oelschlaegel, B. Xhemalce, S. Balasubramanian, S. P. Jackson, *Nat. Chem. Biol.* **2012**, *8*, 301–10.
- [155] W.-C. Huang, T.-Y. Tseng, Y.-T. Chen, C.-C. Chang, Z.-F. Wang, C.-L. Wang, T.-N. Hsu, P.-T. Li, C.-T. Chen, J.-J. Lin, et al., *Nucleic Acids Res.* **2015**, DOI 10.1093/nar/gkv1061.
- [156] A. Laguerre, K. Hukezalie, P. Winckler, F. Katranji, G. Chanteloup, M. Pirrotta, J.-M. Perrier-Cornet, J. M. Y. Wong, D. Monchaud, *J. Am. Chem. Soc.* **2015**, *137*, 8521–5.
- [157] W. Lu, Y. Zhang, D. Liu, Z. Songyang, M. Wan, *Exp. Cell Res.* **2013**, *319*, 133–41.
- [158] T. de Lange, *Genes Dev.* **2005**, *19*, 2100–10.
- [159] W. Traut, M. Szczepanowski, M. Vítková, C. Opitz, F. Marec, J. Zrzavý, *Chromosom. Res.* **2007**, *15*, 371–82.
- [160] M. Mason, A. Schuller, E. Skordalakes, *Curr. Opin. Struct. Biol.* **2011**, *21*, 92–100.
- [161] H. M. R. and R. C. F. James M. Mason, *DNA Replication-Current Advances*, InTech, **2011**.
- [162] T. de Lange, *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 323–9.
- [163] W. Palm, T. de Lange, *Annu. Rev. Genet.* **2008**, *42*, 301–34.
- [164] A. Smogorzewska, B. van Steensel, A. Bianchi, S. Oelmann, M. R. Schaefer, G. Schnapp, T. de Lange, *Mol. Cell. Biol.* **2000**, *20*, 1659–68.
- [165] R. M. Stansel, T. de Lange, J. D. Griffith, *EMBO J.* **2001**, *20*, 5532–40.
- [166] B.-K. Oh, Y.-J. Kim, C. Park, Y. N. Park, *Am. J. Pathol.* **2005**, *166*, 73–80.
- [167] P. Martinez, M. Thanasoula, A. R. Carlos, G. Gómez-López, A. M. Tejera, S. Schoeftner, O. Dominguez, D. G. Pisano, M. Tarsounas, M. A. Blasco, *Nat. Cell Biol.* **2010**, *12*, 768–80.
- [168] K. K. Takai, T. Kibe, J. R. Donigian, D. Frescas, T. de Lange, *Mol. Cell* **2011**, *44*, 647–59.
- [169] S. Ray, J. N. Bandaria, M. H. Qureshi, A. Yildiz, H. Balci, *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 2990–5.
- [170] F. Wang, E. R. Podell, A. J. Zaug, Y. Yang, P. Baciú, T. R. Cech, M. Lei, *Nature* **2007**, *445*, 506–10.
- [171] H. Hwang, N. Buncher, P. L. Opresko, S. Myong, *Structure* **2012**, *20*, 1872–80.
- [172] S. Sandin, D. Rhodes, *Curr. Opin. Struct. Biol.* **2014**, *25*, 104–10.
- [173] H. Gao, R. B. Cervantes, E. K. Mandell, J. H. Otero, V. Lundblad, *Nat. Struct. Mol. Biol.* **2007**, *14*, 208–14.
- [174] L.-Y. Chen, S. Redon, J. Lingner, *Nature* **2012**, *488*, 540–4.
- [175] J. R. Mitchell, K. Collins, *Mol. Cell* **2000**, *6*, 361–371.
- [176] A. Sauerwald, S. Sandin, G. Cristofari, S. H. W. Scheres, J. Lingner, D. Rhodes, *Nat. Struct. Mol. Biol.* **2013**, *20*, 454–60.
- [177] C. W. Greider, E. H. Blackburn, *Cell* **1985**, *43*, 405–13.
- [178] V. Lundblad, J. W. Szostak, *Cell* **1989**, *57*, 633–43.
- [179] S. A. Jacobs, E. R. Podell, T. R. Cech, *Nat. Struct. Mol. Biol.* **2006**, *13*, 218–25.
- [180] Y. Zhao, E. Abreu, J. Kim, G. Stadler, U. Eskiocak, M. P. Terns, R. M. Terns, J. W. Shay, W. E. Wright, *Mol. Cell* **2011**, *42*, 297–307.
- [181] H. Hwang, P. Opresko, S. Myong, *Sci. Rep.* **2014**, *4*, 6391.
- [182] C. B. Harley, H. Vaziri, C. M. Counter, R. C. Allsopp, *Exp. Gerontol.* **27**, 375–82.
- [183] C.-H. Chen, R.-J. Chen, *J. Formos. Med. Assoc.* **2011**, *110*, 275–89.
- [184] C. M. Heaphy, A. P. Subhawong, S.-M. Hong, M. G. Goggins, E. A. Montgomery, E. Gabrielson, G. J. Netto, J. I. Epstein, T. L. Lotan, W. H. Westra, et al., *Am. J. Pathol.* **2011**, *179*, 1608–15.
- [185] L. Oganessian, J. Karlseder, *Mol. Cell* **2011**, *42*, 224–36.
- [186] P. Školáková, S. Foldynová-Trantírková, K. Bednářová, R. Fiala, M. Vorlíčková, L.

- Trantírek, *Nucleic Acids Res.* **2015**, *43*, 4733–45.
- [187] D. Rhodes, H. J. Lipps, *Nucleic Acids Res.* **2015**, *43*, 8627–37.
- [188] M. L. Bochman, K. Paeschke, V. A. Zakian, *Nat. Rev. Genet.* **2012**, *13*, 770–80.
- [189] M. Conacci-Sorrell, L. McFerrin, R. N. Eisenman, *Cold Spring Harb. Perspect. Med.* **2014**, *4*, a014357.
- [190] B. Lüscher, J. Vervoorts, *Gene* **2012**, *494*, 145–160.
- [191] T. Wahlström, M. Arsenian Henriksson, *Biochim. Biophys. Acta* **2015**, *1849*, 563–569.
- [192] E. H. Postel, S. E. Mango, S. J. Flint, *Mol. Cell. Biol.* **1989**, *9*, 5123–33.
- [193] I. Collins, A. Weber, D. Levens, *Mol. Cell. Biol.* **2001**, *21*, 8437–51.
- [194] T. A. Brooks, L. H. Hurley, *Nat. Rev. Cancer* **2009**, *9*, 849–61.
- [195] T. Simonsson, P. Pecinka, M. Kubista, *Nucleic Acids Res.* **1998**, *26*, 1167–72.
- [196] A. T. Phan, Y. S. Modi, D. J. Patel, *J. Am. Chem. Soc.* **2004**, *126*, 8710–6.
- [197] A. T. Phan, V. Kuryavyi, H. Y. Gaw, D. J. Patel, *Nat. Chem. Biol.* **2005**, *1*, 167–73.
- [198] A. Ambrus, D. Chen, J. Dai, R. A. Jones, D. Yang, *Biochemistry* **2005**, *44*, 2048–58.
- [199] J. Seenisamy, E. M. Rezler, T. J. Powell, D. Tye, V. Gokhale, C. S. Joshi, A. Siddiqui-Jain, L. H. Hurley, *J. Am. Chem. Soc.* **2004**, *126*, 8702–8709.
- [200] E. Hatzakis, K. Okamoto, D. Yang, *Biochemistry* **2010**, *49*, 9152–60.
- [201] A. Siddiqui-Jain, C. L. Grand, D. J. Bearss, L. H. Hurley, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 11593–11598.
- [202] L. H. Hurley, D. D. Von Hoff, A. Siddiqui-Jain, D. Yang, *Semin. Oncol.* **2006**, *33*, 498–512.
- [203] T.-M. Ou, Y.-J. Lu, C. Zhang, Z.-S. Huang, X.-D. Wang, J.-H. Tan, Y. Chen, D.-L. Ma, K.-Y. Wong, J. C.-O. Tang, et al., *J. Med. Chem.* **2007**, *50*, 1465–74.
- [204] T. A. Brooks, L. H. Hurley, *Genes Cancer* **2010**, *1*, 641–649.
- [205] G. W. Collie, G. N. Parkinson, *Chem. Soc. Rev.* **2011**, *40*, 5867.
- [206] S. Balasubramanian, L. H. Hurley, S. Neidle, *Nat. Rev. Drug Discov.* **2011**, *10*, 261–75.
- [207] D. Drygin, A. Siddiqui-Jain, S. O'Brien, M. Schwaebe, A. Lin, J. Bliesath, C. B. Ho, C. Proffitt, K. Trent, J. P. Whitten, et al., *Cancer Res.* **2009**, *69*, 7653–61.
- [208] C. Pivetta, L. Lucatello, A. P. Krapcho, B. Gatto, M. Palumbo, C. Sissi, *Bioorg. Med. Chem.* **2008**, *16*, 9331–9.
- [209] J. Seenisamy, S. Bashyam, V. Gokhale, H. Vankayalapati, D. Sun, A. Siddiqui-Jain, N. Streiner, K. Shin-Ya, E. White, W. D. Wilson, et al., *J. Am. Chem. Soc.* **2005**, *127*, 2944–59.
- [210] X. Ji, H. Sun, H. Zhou, J. Xiang, Y. Tang, C. Zhao, *Nucleic Acid Ther.* **2012**, *22*, 127–36.
- [211] J. Wang, K. Lu, S. Xuan, Z. Toh, D. Zhang, F. Shao, *Chem. Commun.* **2013**, *49*, 4758.
- [212] P. Wu, D.-L. Ma, C.-H. Leung, S.-C. Yan, N. Zhu, R. Abagyan, C.-M. Che, *Chemistry* **2009**, *15*, 13008–21.
- [213] S. Maiti, N. K. Chaudhury, S. Chowdhury, *Biochem. Biophys. Res. Commun.* **2003**, *310*, 505–12.
- [214] J. Dash, P. S. Shirude, S.-T. D. Hsu, S. Balasubramanian, *J. Am. Chem. Soc.* **2008**, *130*, 15950–6.
- [215] D.-L. Ma, D. S.-H. Chan, W.-C. Fu, H.-Z. He, H. Yang, S.-C. Yan, C.-H. Leung, *PLoS One* **2012**, *7*, e43278.
- [216] W. J. Chung, B. Heddi, F. Hamon, M.-P. Teulade-Fichou, A. T. Phan, *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 999–1002.
- [217] J. Dai, M. Carver, L. H. Hurley, D. Yang, *J. Am. Chem. Soc.* **2011**, *133*, 17673–17680.
- [218] R. Roberts, D. Govender, *J. Clin. Pathol.* **2015**, jclinpath–2015–203207–.
- [219] D. Wei, G. N. Parkinson, A. P. Reszka, S. Neidle, *Nucleic Acids Res.* **2012**, *40*, 4691–700.
- [220] A. T. Phan, V. Kuryavyi, S. Burge, S. Neidle, D. J. Patel, *J. Am. Chem. Soc.* **2007**, *129*, 4386–92.
- [221] S.-T. D. Hsu, P. Varnai, A. Bugaut, A. P. Reszka, S. Neidle, S. Balasubramanian, *J. Am. Chem. Soc.* **2009**, *131*, 13399–409.
- [222] V. Kuryavyi, A. T. Phan, D. J. Patel, *Nucleic Acids Res.* **2010**, *38*, 6757–73.

Bibliography

- [223] A. Frenzel, F. Grespi, W. Chmielewski, A. Villunger, *Apoptosis* **2009**, *14*, 584–96.
- [224] J. Dai, D. Chen, R. A. Jones, L. H. Hurley, D. Yang, *Nucleic Acids Res.* **2006**, *34*, 5133–44.
- [225] N. Khan, A. Aviñó, R. Tauler, C. González, R. Eritja, R. Gargallo, *Biochimie* **2007**, *89*, 1562–72.
- [226] S. Kendrick, Y. Akiyama, S. M. Hecht, L. H. Hurley, *J. Am. Chem. Soc.* **2009**, *131*, 17667–76.
- [227] S. Kendrick, H.-J. Kang, M. P. Alam, M. M. Madathil, P. Agrawal, V. Gokhale, D. Yang, S. M. Hecht, L. H. Hurley, *J. Am. Chem. Soc.* **2014**, *136*, 4161–71.
- [228] A. M. Duffy, D. J. Bouchier-Hayes, J. H. Harmey, **2000**.
- [229] N. Ferrara, K. Houck, L. Jakeman, D. W. Leung, *Endocr. Rev.* **1992**, *13*, 18–32.
- [230] M. W. Kieran, R. Kalluri, Y.-J. Cho, *Cold Spring Harb. Perspect. Med.* **2012**, *2*, a006593.
- [231] D. Sun, K. Guo, Y.-J. Shin, *Nucleic Acids Res.* **2011**, *39*, 1256–65.
- [232] P. Agrawal, E. Hatzakis, K. Guo, M. Carver, D. Yang, *Nucleic Acids Res.* **2013**, *41*, 10584–92.
- [233] J. I. Bárdos, M. Ashcroft, *Bioessays* **2004**, *26*, 262–9.
- [234] R. De Armond, S. Wood, D. Sun, L. H. Hurley, S. W. Ebbinghaus, *Biochemistry* **2005**, *44*, 16341–50.
- [235] R. G. Ramsay, A. L. Barton, T. J. Gonda, *Expert Opin. Ther. Targets* **2003**, *7*, 235–48.
- [236] S. L. Palumbo, R. M. Memmott, D. J. Uribe, Y. Krotova-Khan, L. H. Hurley, S. W. Ebbinghaus, *Nucleic Acids Res.* **2008**, *36*, 1755–69.
- [237] R. Board, G. C. Jayson, *Drug Resist. Updat.* **2005**, *8*, 75–83.
- [238] Y. Qin, E. M. Rezler, V. Gokhale, D. Sun, L. H. Hurley, *Nucleic Acids Res.* **2007**, *35*, 7698–713.
- [239] Y. Chen, P. Agrawal, R. V Brown, E. Hatzakis, L. Hurley, D. Yang, *J. Am. Chem. Soc.* **2012**, *134*, 13220–3.
- [240] R. Chetty, D. Govender, *J. Clin. Pathol.* **2013**, *66*, 548–50.
- [241] S. Cogoi, M. Paramasivam, B. Spolaore, L. E. Xodo, *Nucleic Acids Res.* **2008**, *36*, 3765–80.
- [242] H. Liu, B. Dibling, B. Spike, A. Dirlam, K. Macleod, *Curr. Opin. Genet. Dev.* **2004**, *14*, 55–64.
- [243] Y. Xu, H. Sugiyama, *Nucleic Acids Res.* **2006**, *34*, 949–54.
- [244] K. W. Lim, L. Lacroix, D. J. E. Yue, J. K. C. Lim, J. M. W. Lim, A. T. Phan, *J. Am. Chem. Soc.* **2010**, *132*, 12331–42.
- [245] E. Besnard, A. Babled, L. Lapasset, O. Milhavet, H. Parrinello, C. Dantec, J.-M. Marin, J.-M. Lemaître, *Nat. Struct. Mol. Biol.* **2012**, *19*, 837–44.
- [246] T. B. C. London, L. J. Barber, G. Mosedale, G. P. Kelly, S. Balasubramanian, I. D. Hickson, S. J. Boulton, K. Hiom, *J. Biol. Chem.* **2008**, *283*, 36132–9.
- [247] K. Paeschke, M. L. Bochman, P. D. Garcia, P. Cejka, K. L. Friedman, S. C. Kowalczykowski, V. A. Zakian, *Nature* **2013**, *497*, 458–62.
- [248] H. Debrauwère, J. Buard, J. Tessier, D. Aubert, G. Vergnaud, A. Nicolas, *Nat. Genet.* **1999**, *23*, 367–71.
- [249] J. Eddy, N. Maizels, *Nucleic Acids Res.* **2008**, *36*, 1321–33.
- [250] M. Trajkovski, M. W. da Silva, J. Plavec, *J. Am. Chem. Soc.* **2012**, *134*, 4132–41.
- [251] D. Hanahan, R. A. Weinberg, *Cell* **2000**, *100*, 57–70.
- [252] T. A. Brooks, S. Kendrick, L. Hurley, *FEBS J.* **2010**, *277*, 3459–69.
- [253] L. H. Hurley, *Nat. Rev. Cancer* **2002**, *2*, 188–200.
- [254] L. Kelland, *Nat. Rev. Cancer* **2007**, *7*, 573–84.
- [255] G. N. Hortobágyi, *Drugs* **1997**, *54 Suppl 4*, 1–7.
- [256] M. P. Knauert, P. M. Glazer, *Hum. Mol. Genet.* **2001**, *10*, 2243–51.
- [257] P. E. Nielsen, *Curr. Pharm. Des.* **2010**, *16*, 3118–23.
- [258] C. Bailly, J. B. Chaires, *Bioconjug. Chem.* **9**, 513–38.

- [259] A. Ali, S. Bhattacharya, *Bioorg. Med. Chem.* **2014**, *22*, 4506–4521.
- [260] W. G. Kaelin, *Nat. Rev. Cancer* **2005**, *5*, 689–98.
- [261] M. Aggarwal, J. A. Sommers, R. H. Shoemaker, R. M. Brosh, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 1525–30.
- [262] K. I. E. McLuckie, M. Di Antonio, H. Zecchini, J. Xian, C. Caldas, B.-F. Krippendorff, D. Tannahill, C. Lowe, S. Balasubramanian, *J. Am. Chem. Soc.* **2013**, *135*, 9640–3.
- [263] S. Balasubramanian, S. Neidle, *Curr. Opin. Chem. Biol.* **2009**, *13*, 345–353.
- [264] S. A. Ohnmacht, S. Neidle, *Bioorg. Med. Chem. Lett.* **2014**, *24*, 2602–12.
- [265] V. Sekaran, J. Soares, M. B. Jarstfer, *J. Med. Chem.* **2014**, *57*, 521–38.
- [266] A. M. Zahler, J. R. Williamson, T. R. Cech, D. M. Prescott, *Nature* **1991**, *350*, 718–20.
- [267] D. Sun, B. Thompson, B. E. Cathers, M. Salazar, S. M. Kerwin, J. O. Trent, T. C. Jenkins, S. Neidle, L. H. Hurley, *J. Med. Chem.* **1997**, *40*, 2113–6.
- [268] M. Gunaratnam, O. Greciano, C. Martins, A. P. Reszka, C. M. Schultes, H. Morjani, J.-F. Riou, S. Neidle, *Biochem. Pharmacol.* **2007**, *74*, 679–89.
- [269] E. Salvati, C. Leonetti, A. Rizzo, M. Scarsella, M. Mottolese, R. Galati, I. Sperduti, M. F. G. Stevens, M. D'Incalci, M. Blasco, et al., *J. Clin. Invest.* **2007**, *117*, 3236–47.
- [270] L. Kelland, *Clin. Cancer Res.* **2007**, *13*, 4960–4963.
- [271] S. Neidle, *FEBS J.* **2010**, *277*, 1118–25.
- [272] A. N. Lane, J. B. Chaires, R. D. Gray, J. O. Trent, *Nucleic Acids Res.* **2008**, *36*, 5482–515.
- [273] A. N. Lane, *Biochimie* **2012**, *94*, 277–86.
- [274] J. Liu, F. Kouzine, Z. Nie, H.-J. Chung, Z. Elisha-Feil, A. Weber, K. Zhao, D. Levens, *EMBO J.* **2006**, *25*, 2119–30.
- [275] F. Kouzine, S. Sanford, Z. Elisha-Feil, D. Levens, *Nat. Struct. Mol. Biol.* **2008**, *15*, 146–154.
- [276] V. González, L. H. Hurley, *Annu. Rev. Pharmacol. Toxicol.* **2010**, *50*, 111–29.
- [277] P. Armas, S. Nasif, N. B. Calcaterra, *J. Cell. Biochem.* **2008**, *103*, 1013–36.
- [278] T. Tomonaga, D. Levens, *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 5830–5.
- [279] V. González, K. Guo, L. Hurley, D. Sun, *J. Biol. Chem.* **2009**, *284*, 23622–35.
- [280] A. Siddiqui-Jain, C. L. Grand, D. J. Bearss, L. H. Hurley, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 11593–8.
- [281] A. Bugaut, S. Balasubramanian, *Nucleic Acids Res.* **2012**, *40*, 4727–41.
- [282] C. Cheong, P. B. Moore, *Biochemistry* **1992**, *31*, 8406–14.
- [283] G. Biffi, M. Di Antonio, D. Tannahill, S. Balasubramanian, *Nat. Chem.* **2014**, *6*, 75–80.
- [284] J. L. Huppert, A. Bugaut, S. Kumari, S. Balasubramanian, *Nucleic Acids Res.* **2008**, *36*, 6260–8.
- [285] C. M. Azzalin, P. Reichenbach, L. Khoriantuli, E. Giulotto, J. Lingner, *Science* **2007**, *318*, 798–801.
- [286] R. Arora, C. M. C. Brun, C. M. Azzalin, *Prog. Mol. Subcell. Biol.* **2011**, *51*, 65–94.
- [287] B. Luke, J. Lingner, *EMBO J.* **2009**, *28*, 2503–10.
- [288] H. Martadinata, A. T. Phan, *J. Am. Chem. Soc.* **2009**, *131*, 2570–8.
- [289] H. Martadinata, A. T. Phan, *Biochemistry* **2013**, *52*, 2176–83.
- [290] G. W. Collie, S. M. Haider, S. Neidle, G. N. Parkinson, *Nucleic Acids Res.* **2010**, *38*, 5569–80.
- [291] G. W. Collie, S. Sparapani, G. N. Parkinson, S. Neidle, *J. Am. Chem. Soc.* **2011**, *133*, 2721–8.
- [292] D.-H. Zhang, T. Fujimoto, S. Saxena, H.-Q. Yu, D. Miyoshi, N. Sugimoto, *Biochemistry* **2010**, *49*, 4554–63.
- [293] A. Arora, S. Maiti, *J. Phys. Chem. B* **2009**, *113*, 10515–20.
- [294] M. Di Antonio, G. Biffi, A. Mariani, E.-A. Raiber, R. Rodriguez, S. Balasubramanian, *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 11073–8.
- [295] A. Arora, M. Dutkiewicz, V. Scaria, M. Hariharan, S. Maiti, J. Kurreck, *RNA* **2008**, *14*, 1290–6.

Bibliography

- [296] S. Lammich, F. Kamp, J. Wagner, B. Nuscher, S. Zilow, A.-K. Ludwig, M. Willem, C. Haass, *J. Biol. Chem.* **2011**, *286*, 45063–72.
- [297] S. Kumari, A. Bugaut, S. Balasubramanian, *Biochemistry* **2008**, *47*, 12664–9.
- [298] D. Bhattacharyya, P. Diamond, S. Basu, *Biochemistry* **2015**, *54*, 1879–85.
- [299] M. J. Morris, Y. Negishi, C. Pázsint, J. D. Schonhofs, S. Basu, *J. Am. Chem. Soc.* **2010**, *132*, 17831–9.
- [300] K. Halder, M. Wieland, J. S. Hartig, *Nucleic Acids Res.* **2009**, *37*, 6811–7.
- [301] M. Webba da Silva, *Methods* **2007**, *43*, 264–77.
- [302] M. Adrian, B. Heddi, A. T. Phan, *Methods* **2012**, *57*, 11–24.
- [303] D. Miyoshi, H. Karimata, N. Sugimoto, *Nucleosides. Nucleotides Nucleic Acids* **2007**, *26*, 589–95.
- [304] K. N. Luu, A. T. Phan, V. Kuryavyi, L. Lacroix, D. J. Patel, *J. Am. Chem. Soc.* **2006**, *128*, 9963–70.
- [305] A. Karsiotis, M. Webba da Silva, *Molecules* **2012**, *17*, 13073–13086.
- [306] C. J. Lech, J. K. Cheow Lim, J. M. Wen Lim, S. Amrane, B. Heddi, A. T. Phan, *Biophys. J.* **2011**, *101*, 1987–98.
- [307] V. Singh, A. Benz, J. S. Hartig, *Chemistry* **2011**, *17*, 10838–43.
- [308] V. V. Cheong, B. Heddi, C. J. Lech, A. T. Phan, *Nucleic Acids Res.* **2015**, gkv826–.
- [309] C. G. Peng, M. J. Damha, *Nucleic Acids Res.* **2007**, *35*, 4977–88.
- [310] C. J. Lech, Z. Li, B. Heddi, A. T. Phan, *Chem. Commun. (Camb)*. **2012**, *48*, 11425–7.
- [311] N. Martín-Pintado, M. Yahyaee-Anzahae, G. F. Deleavey, G. Portella, M. Orozco, M. J. Damha, C. González, *J. Am. Chem. Soc.* **2013**, *135*, 5344–7.
- [312] J. F. Trempe, C. J. Wilds, A. Y. Denisov, R. T. Pon, M. J. Damha, K. Gehring, *J. Am. Chem. Soc.* **2001**, *123*, 4896–903.
- [313] Z. Li, C. J. Lech, A. T. Phan, *Nucleic Acids Res.* **2014**, *42*, 4068–79.
- [314] M. Marušič, R. N. Veedu, J. Wengel, J. Plavec, *Nucleic Acids Res.* **2013**, *41*, 9524–36.
- [315] J. T. Nielsen, K. Arar, M. Petersen, *Nucleic Acids Res.* **2006**, *34*, 2006–14.
- [316] I. Russo Krauss, G. N. Parkinson, A. Merlino, C. A. Mattia, A. Randazzo, E. Novellino, L. Mazzarella, F. Sica, *Acta Crystallogr. D. Biol. Crystallogr.* **2014**, *70*, 362–70.
- [317] V. . Sklenar, A. Bax, *J. Magn. Res.* **1987**, *74*, 469.
- [318] K. L. Greene, Y. Wang, D. Live, *J. Biomol. NMR* **1995**, *5*, 333–8.
- [319] J. Ippel, S. Wijmenga, R. deJong, **1996**, *34*.
- [320] Z. Vokáčová, F. M. Bickelhaupt, J. Sponer, V. Sychrovský, *J. Phys. Chem. A* **2009**, *113*, 8379–86.
- [321] H. Schwalbe, J. P. Marino, G. C. King, R. Wechselberger, W. Bermel, C. Griesinger, *J. Biomol. NMR* **1994**, *4*, 631–44.
- [322] S. S. Wijmenga, B. N. M. van Buuren, *Prog. Nucl. Magn. Reson. Spectrosc.* **1998**, *32*, 287–387.
- [323] M. M. Mooren, S. S. Wijmenga, G. A. van der Marel, J. H. van Boom, C. W. Hilbers, *Nucleic Acids Res.* **1994**, *22*, 2658–66.
- [324] C. A. G. Haasnoot, F. A. A. M. de Leeuw, C. Altona, *Tetrahedron* **1980**, *36*, 2783–2792.
- [325] Kurt Wüthrich, *NMR of Proteins and Nucleic Acids*, New York, **1986**.
- [326] A. T. Phan, *J. Biomol. NMR* **2000**, *16*, 175–8.
- [327] A. J. Dingley, J. E. Masse, J. Feigon, S. Grzesiek, *J. Biomol. NMR* **2000**, *16*, 279–89.
- [328] A. Majumdar, Y. Gosser, D. J. Patel, *J. Biomol. NMR* **2001**, *21*, 289–306.
- [329] A. T. Phan, V. Kuryavyi, J. C. Darnell, A. Serganov, A. Majumdar, S. Ilin, T. Raslin, A. Polonskaia, C. Chen, D. Clain, et al., *Nat. Struct. Mol. Biol.* **2011**, *18*, 796–804.
- [330] J. F. Milligan, D. R. Groebe, G. W. Witherell, O. C. Uhlenbeck, *Nucleic Acids Res.* **1987**, *15*, 8783–98.
- [331] D. P. Zimmer, D. M. Crothers, *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 3091–3095.
- [332] W. J. Chung, B. Heddi, M. Tera, K. Iida, K. Nagasawa, A. T. Phan, *J. Am. Chem. Soc.* **2013**, *135*, 13495–501.

- [333] J. M. Louis, R. G. Martin, G. M. Clore, A. M. Gronenborn, *J. Biol. Chem.* **1998**, *273*, 2374–8.
- [334] F. H. T. Nelissen, F. C. Girard, M. Tessari, H. A. Heus, S. S. Wijmenga, *Nucleic Acids Res.* **2009**, *37*, e114–e114.
- [335] C. Kojima, A. Ono, M. Kainosho, *Methods Enzymol.* **2001**, *338*, 261–83.
- [336] A. T. Phan, D. J. Patel, *J. Am. Chem. Soc.* **2002**, *124*, 1160–1.
- [337] M. Marusic, P. Sket, L. Bauer, V. Viglasky, J. Plavec, *Nucleic Acids Res.* **2012**, *40*, 6946–56.
- [338] A. L. Breeze, *Prog. Nucl. Magn. Reson. Spectrosc.* **2000**, *36*, 323–372.
- [339] A. J. Dingley, J. E. Masse, R. D. Peterson, M. Barfield, J. Feigon, S. Grzesiek, *J. Am. Chem. Soc.* **1999**, *121*, 6019–6027.
- [340] N. V. Hud, P. Schultze, V. Sklenár, J. Feigon, *J. Mol. Biol.* **1999**, *285*, 233–43.
- [341] F. W. Smith, J. Feigon, *Biochemistry* **1993**, *32*, 8682–8692.
- [342] S. Haider, G. N. Parkinson, S. Neidle, *J. Mol. Biol.* **2002**, *320*, 189–200.
- [343] S. Basu, A. A. Szewczak, M. Cocco, S. A. Strobel, *J. Am. Chem. Soc.* **2000**, *122*, 3240–3241.
- [344] H. Deng, W. H. Braunlin, *J. Mol. Biol.* **1996**, *255*, 476–83.
- [345] A. Wong, R. Ida, G. Wu, *Biochem. Biophys. Res. Commun.* **2005**, *337*, 363–6.
- [346] R. Ida, G. Wu, *J. Am. Chem. Soc.* **2008**, *130*, 3590–3602.
- [347] D. Rovnyak, M. Baldus, G. Wu, N. V. Hud, J. Feigon, R. G. Griffin, *J. Am. Chem. Soc.* **2000**, *122*, 11423–11429.
- [348] G. Wu, A. Wong, *Biochem. Biophys. Res. Commun.* **2004**, *323*, 1139–44.
- [349] G. Wu, A. Wong, Z. Gan, J. T. Davis, *J. Am. Chem. Soc.* **2003**, *125*, 7182–3.
- [350] I. C. M. Kwan, X. Mo, G. Wu, *J. Am. Chem. Soc.* **2007**, *129*, 2398–407.
- [351] J. R. Bothe, E. N. Nikolova, C. D. Eichhorn, J. Chugh, A. L. Hansen, H. M. Al-Hashimi, *Nat. Methods* **2011**, *8*, 919–31.
- [352] J. Stehle, R. Silvers, K. Werner, D. Chatterjee, S. Gande, F. Scholz, A. Dutta, J. Wachtveitl, J. Klein-Seetharaman, H. Schwalbe, *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 2078–84.
- [353] J. Rinnenthal, D. Wagner, T. Marquardsen, A. Krahn, F. Engelke, H. Schwalbe, *J. Magn. Reson.* **2015**, *251*, 84–93.
- [354] A. L. Lieblein, J. Buck, K. Schlepckow, B. Fürtig, H. Schwalbe, *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 250–3.
- [355] K. H. Mok, T. Nagashima, I. J. Day, J. A. Jones, C. J. V. Jones, C. M. Dobson, P. J. Hore, *J. Am. Chem. Soc.* **2003**, *125*, 12484–92.
- [356] M.-K. Lee, M. Gal, L. Frydman, G. Varani, *Proc. Natl. Acad. Sci.* **2010**, *107*, 9192–9197.
- [357] E. Rennella, B. Brutscher, *Chemphyschem* **2013**, *14*, 3059–70.
- [358] D. Monchaud, M. P. Teulade-Fichou, *Org. Biomol. Chem.* **2008**, *6*, 627–636.
- [359] T. M. Ou, Y. J. Lu, J. H. Tan, Z. S. Huang, K. Y. Wong, L. Q. Gu, *ChemMedChem* **2008**, *3*, 690–713.
- [360] A. De Cian, L. Lacroix, C. Douarre, N. Temime-Smaali, C. Trentesaux, J. F. Riou, J. L. Mergny, *Biochimie* **2008**, *90*, 131–155.
- [361] P. Murat, Y. Singh, E. Defrancq, *Chem. Soc. Rev.* **2011**, *40*, 5293–307.
- [362] Q. Li, J.-F. Xiang, Q.-F. Yang, H.-X. Sun, A.-J. Guan, Y.-L. Tang, *Nucleic Acids Res.* **2013**, *41*, D1115–23.
- [363] L. Martino, A. Virno, B. Pagano, A. Virgilio, S. Di Micco, A. Galeone, C. Giancola, G. Bifulco, L. Mayol, A. Randazzo, *J. Am. Chem. Soc.* **2007**, *129*, 16048–16056.
- [364] Q. Li, J. Xiang, X. Li, L. Chen, X. Xu, Y. Tang, Q. Zhou, L. Li, H. Zhang, H. Sun, et al., *Biochimie* **2009**, *91*, 811–9.
- [365] S. Cosconati, L. Marinelli, R. Trotta, A. Virno, S. De Tito, R. Romagnoli, B. Pagano, V. Limongelli, C. Giancola, P. G. Baraldi, et al., *J. Am. Chem. Soc.* **2010**, *132*, 6425–6433.
- [366] A. M. Burger, F. P. Dai, C. M. Schultes, A. P. Reszka, M. J. Moore, J. A. Double, S. Neidle,

- Cancer Res.* **2005**, *65*, 1489–1496.
- [367] N. H. . P. G. N. . R. A. P. . N. S. Campbell, *J. Am. Chem. Soc.* **2008**, *130*, 6722–6724.
- [368] S. M. Gowan, R. Heald, M. F. G. Stevens, L. R. Kelland, *Mol. Pharmacol.* **2001**, *60*, 981–988.
- [369] E. Gavathiotis, R. . Heald, M. F. . Stevens, M. S. Searle, *J. Mol. Biol.* **2003**, *334*, 25–36.
- [370] W. Duan, A. Rangan, H. Vankayalapati, M. Y. Kim, Q. Zeng, D. Sun, H. Han, O. Y. Fedoroff, D. Nishioka, S. Y. Rha, et al., *Mol. Cancer Ther.* **2001**, *1*, 103–20.
- [371] C. Bazzicalupi, M. Ferraroni, A. . Bilia, F. Scheggi, P. Gratteri, *Nucleic Acids Res.* **2013**, *41*, 632–638.
- [372] F. Cuenca, O. Greciano, M. Gunaratnam, S. Haider, D. Munnur, R. Nanjunda, W. D. Wilson, S. Neidle, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 1668–73.
- [373] G. N. Parkinson, F. Cuenca, S. Neidle, *J. Mol. Biol.* **2008**, *381*, 1145–1156.
- [374] G. W. Collie, R. Promontorio, S. M. Hampel, M. Micco, S. Neidle, G. N. Parkinson, *J. Am. Chem. Soc.* **2012**, *134*, 2723–31.
- [375] M. Micco, G. W. Collie, A. G. Dale, S. A. Ohnmacht, I. Pazitna, M. Gunaratnam, A. P. Reszka, S. Neidle, *J. Med. Chem.* **2013**, *56*, 2959–74.
- [376] O. Y. Fedoroff, M. Salazar, H. Han, V. V Chemeris, S. M. Kerwin, L. H. Hurley, *Biochemistry* **1998**, *37*, 12367–74.
- [377] J. T. Kern, P. W. Thomas, S. M. Kerwin, *Biochemistry* **2002**, *41*, 11379–89.
- [378] A. De Cian, E. Delemos, J.-L. Mergny, M.-P. Teulade-Fichou, D. Monchaud, *J. Am. Chem. Soc.* **2007**, *129*, 1856–7.
- [379] C.-C. Chang, J.-Y. Wu, C.-W. Chien, W.-S. Wu, H. Liu, C.-C. Kang, L.-J. Yu, T.-C. Chang, *Anal. Chem.* **2003**, *75*, 6177–83.
- [380] I. Lubitz, N. Borovok, A. Kotlyar, *Biochemistry* **2007**, *46*, 12925–9.
- [381] J. M. Nicoludis, S. P. Barrett, J.-L. Mergny, L. A. Yatsunyk, *Nucleic Acids Res.* **2012**, *40*, 5432–47.
- [382] J. M. Nicoludis, S. T. Miller, P. D. Jeffrey, S. P. Barrett, P. R. Rablen, T. J. Lawton, L. A. Yatsunyk, *J. Am. Chem. Soc.* **2012**, *134*, 20446–56.
- [383] R. Rodriguez, S. Müller, J. A. Yeoman, C. Trentesaux, J.-F. Riou, S. Balasubramanian, *J. Am. Chem. Soc.* **2008**, *130*, 15758–9.
- [384] R. Rodriguez, K. M. Miller, J. V Forment, C. R. Bradshaw, M. Nikan, S. Britton, T. Oelschlaegel, B. Xhemalce, S. Balasubramanian, S. P. Jackson, *Nat. Chem. Biol.* **2012**, *8*, 301–10.
- [385] K. Shin-ya, K. Wierzba, K. Matsuo, T. Ohtani, Y. Yamada, K. Furihata, Y. Hayakawa, H. Seto, *J. Am. Chem. Soc.* **2001**, *123*, 1262–1263.
- [386] D. Monchaud, A. Granzhan, N. Saettel, A. Guédin, J.-L. Mergny, M.-P. Teulade-Fichou, *J. Nucleic Acids* **2010**, *2010*, DOI 10.4061/2010/525862.
- [387] L. Stefan, A. Guédin, S. Amrane, N. Smith, F. Denat, J.-L. Mergny, D. Monchaud, *Chem. Commun. (Camb)*. **2011**, *47*, 4992–4.
- [388] A. Laguerre, L. Stefan, M. Larrouy, D. Genest, J. Novotna, M. Pirrotta, D. Monchaud, *J. Am. Chem. Soc.* **2014**, *136*, 12406–12414.
- [389] Y. P. Kumar, S. Bhowmik, R. N. Das, I. Bessi, S. Paladhi, R. Ghosh, H. Schwalbe, J. Dash, *Chemistry* **2013**, *19*, 11502–6.
- [390] Y. P. Kumar, R. N. Das, S. Kumar, O. M. Schütte, C. Steinem, J. Dash, *Chemistry* **2014**, *20*, 3023–8.
- [391] A. De Cian, G. Cristofari, P. Reichenbach, E. De Lemos, D. Monchaud, M. P. Teulade-Fichou, K. Shin-Ya, L. Lacroix, J. Lingner, J. L. Mergny, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 17347–17352.
- [392] S. Søndergaard, M. Aznauryan, E. K. Hastrup, B. Schiøtt, V. Birkedal, B. Corry, *Chemphyschem* **2015**, *16*, 2562–70.
- [393] M. Micco, G. . Collie, A. . Dale, S. . Ohnmacht, I. Pazitna, M. Gunaratnam, A. . Reszka, S. Neidle, *J. Med. Chem.* **2013**, *56*, 2959–2974.

- [394] G. W. Collie, N. H. Campbell, S. Neidle, *Nucleic Acids Res.* **2015**, *43*, 4785–4799.
- [395] J. . Nicoludis, S. . Miller, P. . Jeffrey, S. . Barrett, P. . Rablen, T. . Lawton, L. A. Yatsunyk, *J. Am. Chem. Soc.* **2012**, *134*, 20446–20456.
- [396] A. Cummaro, I. Fotticchia, M. Franceschin, C. Giancola, L. Petraccone, *Biochimie* **2011**, *93*, 1392–400.
- [397] S. Neidle, *Therapeutic Applications of Quadruplex Nucleic Acids*, Elsevier, **2012**.
- [398] V. Gabelica, *Biochimie* **2014**, *105*, 1–3.
- [399] R. D. Gray, J. B. Chaires, *Nucleic Acids Res.* **2008**, *36*, 4191–203.
- [400] R. D. Gray, J. O. Trent, J. B. Chaires, *J. Mol. Biol.* **2014**, *426*, 1629–50.
- [401] A. I. Karsisiotis, N. M. Hessari, E. Novellino, G. P. Spada, A. Randazzo, M. Webba da Silva, *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 10645–8.
- [402] D. Thirumalai, C. Hyeon, *Biochemistry* **2005**, *44*, 4957–70.
- [403] C. K. Kwok, M. E. Sherlock, P. C. Bevilacqua, *Angew. Chem. Int. Ed. Engl.* **2013**, *52*, 683–6.
- [404] P. Stadlbauer, M. Krepl, T. E. Cheatham, J. Koca, J. Sponer, *Nucleic Acids Res.* **2013**, *41*, 7128–43.
- [405] P. Stadlbauer, L. Trantírek, T. E. Cheatham, J. Koča, J. Sponer, *Biochimie* **2014**, *105*, 22–35.
- [406] P. Stadlbauer, P. Kührová, P. Banáš, J. Koča, G. Bussi, L. Trantírek, M. Otyepka, J. Šponer, *Nucleic Acids Res.* **2015**, *43*, 9626–44.
- [407] B. Islam, P. Stadlbauer, M. Krepl, J. Koca, S. Neidle, S. Haider, J. Sponer, *Nucleic Acids Res.* **2015**, *43*, 8673–93.
- [408] D. Koirala, T. Mashimo, Y. Sannohe, Z. Yu, H. Mao, H. Sugiyama, *Chem. Commun. (Camb)*. **2012**, *48*, 2006–8.
- [409] T. Mashimo, H. Yagi, Y. Sannohe, A. Rajendran, H. Sugiyama, *J. Am. Chem. Soc.* **2010**, *132*, 14910–8.
- [410] A. Rajendran, M. Endo, K. Hidaka, H. Sugiyama, *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 4107–12.
- [411] S. Čeru, P. Šket, I. Prislán, J. Lah, J. Plavec, *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 4881–4.
- [412] J. Y. Lee, B. Okumus, D. S. Kim, T. Ha, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 18938–43.
- [413] S. L. Noer, S. Preus, D. Gudnason, M. Aznauryan, J.-L. Mergny, V. Birkedal, *Nucleic Acids Res.* **2015**, DOI 10.1093/nar/gkv1320.
- [414] A. Heckel, M. Famulok, *Biochimie* **2008**, *90*, 1096–107.
- [415] N. C. Seeman, *Annu. Rev. Biochem.* **2010**, *79*, 65–87.
- [416] Y. Lu, J. Liu, *Curr. Opin. Biotechnol.* **2006**, *17*, 580–8.
- [417] M. R. Jones, N. C. Seeman, C. A. Mirkin, *Science* **2015**, *347*, 1260901.
- [418] Z. Li, C. A. Mirkin, *J. Am. Chem. Soc.* **2005**, *127*, 11568–9.
- [419] D. P. N. Gonçalves, T. L. Schmidt, M. B. Koeppel, A. Heckel, *Small* **2010**, *6*, 1347–52.
- [420] Y. Wen, L. Xu, C. Li, H. Du, L. Chen, B. Su, Z. Zhang, X. Zhang, Y. Song, *Chem. Commun. (Camb)*. **2012**, *48*, 8410–2.
- [421] T. Li, D. Ackermann, A. M. Hall, M. Famulok, *J. Am. Chem. Soc.* **2012**, *134*, 3508–16.
- [422] J. Elbaz, M. Moshe, B. Shlyahovsky, I. Willner, *Chemistry* **2009**, *15*, 3411–8.
- [423] F. Wang, X. Liu, I. Willner, *Angew. Chem. Int. Ed. Engl.* **2015**, *54*, 1098–129.
- [424] J. Li, X. Wang, X. Liang, *Chem. Asian J.* **2014**, *9*, 3344–58.
- [425] W. Szymański, J. M. Beierle, H. A. V Kistemaker, W. A. Velema, B. L. Feringa, *Chem. Rev.* **2013**, *113*, 6114–78.
- [426] S. Rehm, M. O. Lenz, S. Mensch, H. Schwalbe, J. Wachtveitl, *Chem. Phys.* **2006**, *323*, 28–35.
- [427] A. A. Beharry, G. A. Woolley, *Chem. Soc. Rev.* **2011**, *40*, 4422–37.
- [428] L. Wu, K. Koumoto, N. Sugimoto, *Chem. Commun. (Camb)*. **2009**, 1915–7.

- [429] C.-H. Lu, I. Willner, *Angew. Chem. Int. Ed. Engl.* **2015**, *54*, 12212–35.
- [430] M. McCullagh, I. Franco, M. A. Ratner, G. C. Schatz, *J. Am. Chem. Soc.* **2011**, *133*, 3452–9.

Complete list of publications

1. "Small Molecule Regulated Dynamic Structural changes of Human G-quadruplexes." Manish Debnath, Shirsendu Ghosh, Deepanjan Panda, Irene Bessi, Harald Schwalbe, Kankan Bhattacharyya, Jyotirmayee Dash. *Chemical Science*, accepted.
2. "Fluorescent Dansyl-guanosine Conjugates That Bind *c-MYC* Promoter G-quadruplex and Down-regulate *c-MYC* Expression." Y. Pavan Kumar, Puja Saha, Dhurjhoti Saha, Irene Bessi, Harald Schwalbe, Shantanu Chowdhury, Jyotirmayee Dash. *Chembiochem.*, in press.
3. "Photoresponsive formation of an intermolecular minimal G-quadruplex motif." Julie Thevarpadam,[#] Irene Bessi,[#] Oliver Binas,[#] Diana P. N. Gonçalves, Chavdar Slavov, Hendrick R. A. Jonker, Christian Richter, Joseph Wachtveitl, Harald Schwalbe, Alexander Heckel.[#] Equal contribution. *Angew. Chem. Int. Ed. Engl.*, in press.
4. "A Nucleus-Imaging Probe That Selectively Stabilizes a Minor Conformation of *c-MYC* G-quadruplex and Down-regulates *c-MYC* Transcription in Human Cancer Cells." Deepanjan Panda, Manish Debnath, Samir Mandal, Irene Bessi, Harald Schwalbe, Jyotirmayee Dash. *Sci. Rep.* Volume 19, Issue 5, page 13183 (2015).
5. "Involvement of long-lived intermediate states in the complex folding pathway of the human telomeric G-quadruplex." Irene Bessi, Hendrick R. A. Jonker, Christian Richter, Harald Schwalbe. *Angew. Chem. Int. Ed. Engl.*, Volume 54, Issue 29, pages 8444-8 (2015).
6. "Noncovalent spin labeling of riboswitch RNAs to obtain long-range structural NMR restraints." Christina Helmling, Irene Bessi, Anna Wacker, Kai A. Schnorr, Hendrick R. A. Jonker, Christian Richter, Dominic Wagner, Michael Kreibich, Harald Schwalbe. *ACS Chem Biol.*, Volume 9, Issue 6, pages 1330-9 (2014).
7. "A Fluorescent guanosine dinucleoside as a selective switch-on sensor for *c-myc* G-quadruplex DNA with potent anticancer activities." Y. Pavan Kumar, Sudipta Bhowmik, Rabindra N. Das, Irene Bessi, Sushovan Paladhi, Rita Ghosh, Harald Schwalbe, Jyotirmayee Dash. *Chem. Eur. J.*, Volume 19, Issue 35, pages 11502–11506 (2013).
Article featured as Inside Cover in *Chem. Eur. J.* 35/2013.
8. "Spectroscopic, molecular modeling, and NMR-spectroscopic investigation of the binding mode of the natural alkaloids berberine and sanguinarine to human telomeric G-quadruplex DNA." Irene Bessi, Carla Bazzicalupi, Christian Richter, Hendrick R. A. Jonker, Krishna Saxena, Claudia Sissi, Matteo Chioccioli, Sara Bianco, Anna Rita Bilia, Harald Schwalbe, Paola Gratter. *ACS Chem Biol.*, Volume 7, Issue 6, pages 1109-19 (2012).