



# Role of Alkali Metal Ions in G-Quadruplex Nucleic Acid Structure and Stability

Eric Largy, Jean-Louis Mergny, Valérie Gabelica

## ► To cite this version:

Eric Largy, Jean-Louis Mergny, Valérie Gabelica. Role of Alkali Metal Ions in G-Quadruplex Nucleic Acid Structure and Stability. The Alkali Metal Ions: Their Role for Life, Springer Verlag (Germany), pp.203-258, 2016, 10.1007/978-3-319-21756-7\_7 . hal-01524114

**HAL Id: hal-01524114**

**<https://hal.science/hal-01524114>**

Submitted on 17 May 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## The Role of Alkali Metal Ions in G-Quadruplex Nucleic Acid Structure and Stability

*Eric Largy,<sup>1,2</sup> Jean-Louis Mergny,<sup>1,2</sup> and Valérie Gabelica<sup>1,2</sup>*

<sup>1</sup>Université Bordeaux, IECB, ARNA Laboratory, F-33600 Pessac, France

<sup>2</sup>INSERM, U869, ARNA Laboratory, F-33000 Bordeaux, France

<eric.largy@inserm.fr>

<jean-louis.mergny@inserm.fr>

<valerie.gabelica@inserm.fr>

### Contents

#### ABSTRACT

- 1 INTRODUCTION: G-QUADRUPLEX NUCLEIC ACIDS
  - 1.1 Overview of Structure
  - 1.2 Stabilizing Interactions
  - 1.3 Alkali Metal Ion Coordination in G-Quartets
    - 1.3.1 Cation Preference
    - 1.3.2 Cation Binding Energetics
    - 1.3.3 Cation Location
- 2 METHODS TO STUDY G-QUADRUPLEX NUCLEIC ACIDS
  - 2.1 Folding Topology
    - 2.1.1 X-Ray Crystallography
    - 2.1.2 Nuclear Magnetic Resonance Spectroscopy
    - 2.1.3 Molecular Modeling
    - 2.1.4 Electronic Circular Dichroism Spectroscopy
    - 2.1.5 UV Absorption Spectroscopy
    - 2.1.6 Separative Techniques
    - 2.1.7 Native Mass Spectrometry
    - 2.1.8 Miscellaneous
  - 2.2 Cation Coordination
    - 2.2.1 X-Ray Crystallography
    - 2.2.2 Nuclear Magnetic Resonance Spectroscopy
    - 2.2.3 Native Mass Spectrometry
- 3 ROLE OF ALKALI METAL IONS IN G-QUADRUPLEX STABILITY
  - 3.1 Case Study: dTG<sub>3-5</sub>T Tetramolecular G-Quadruplexes
  - 3.2 General Trends
    - 3.2.1 Libraries
    - 3.2.2 Human Telomeric Sequences
    - 3.2.3 Other Sequences and Overview

3.2.4	Summary
3.3	Kinetics of Strand Association, Dissociation, Folding and Unfolding
3.4	Cation Exchange Mechanisms
4	INFLUENCE OF ALKALI METAL IONS ON G-QUADRUPLEX STRUCTURES
4.1	Case Study: The Human Telomeric G-Quadruplex Sequence
4.1.1	The Intramolecular Folding of dAGGG(TTAGGG) <sub>3</sub>
4.1.2	Other Human Telomeric Sequences
4.2	Other Sequences
4.3	General Trends
5	CATION-DEPENDENT CONFORMATIONAL SWITCHING
6	CONCLUDING REMARKS AND FUTURE DIRECTIONS
	ABBREVIATIONS AND DEFINITIONS
	ACKNOWLEDGMENTS
	REFERENCES

**Abstract** G-quadruplexes are guanine-rich nucleic acids that fold by forming successive quartets of guanines (the G-tetrads), stabilized by intra-quartet hydrogen bonds, inter-quartet stacking, and cation coordination. This specific although highly polymorphic type of secondary structure deviates significantly from the classical B-DNA duplex. G-quadruplexes are detectable in human cells and are strongly suspected to be involved in a number of biological processes at the DNA and RNA levels. The vast structural polymorphism exhibited by G-quadruplexes, together with their putative biological relevance, makes them attractive therapeutic targets compared to canonical duplex DNA. This chapter focuses on the essential and specific coordination of alkali metal cations by G-quadruplex nucleic acids, and most notably on studies highlighting cation-dependent dissimilarities in their stability, structure, formation, and interconversion. Section 1 surveys G-quadruplex structures and their interactions with alkali metal ions while section 2 presents analytical methods used to study G-quadruplexes. The influence of alkali cations on the stability, structure, and kinetics of formation of G-quadruplex structures of quadruplexes will be discussed in sections 3 and 4. Section 5 focuses on the cation-induced interconversion of G-quadruplex structures. In sections 3 to 5, we will particularly emphasize the comparisons between cations, most often K<sup>+</sup> and Na<sup>+</sup> because of their prevalence in the literature and in cells.

**Keywords** DNA · folding · G-quadruplex · G-quartet · interconversion · metal ions · methods · RNA · stability · structure

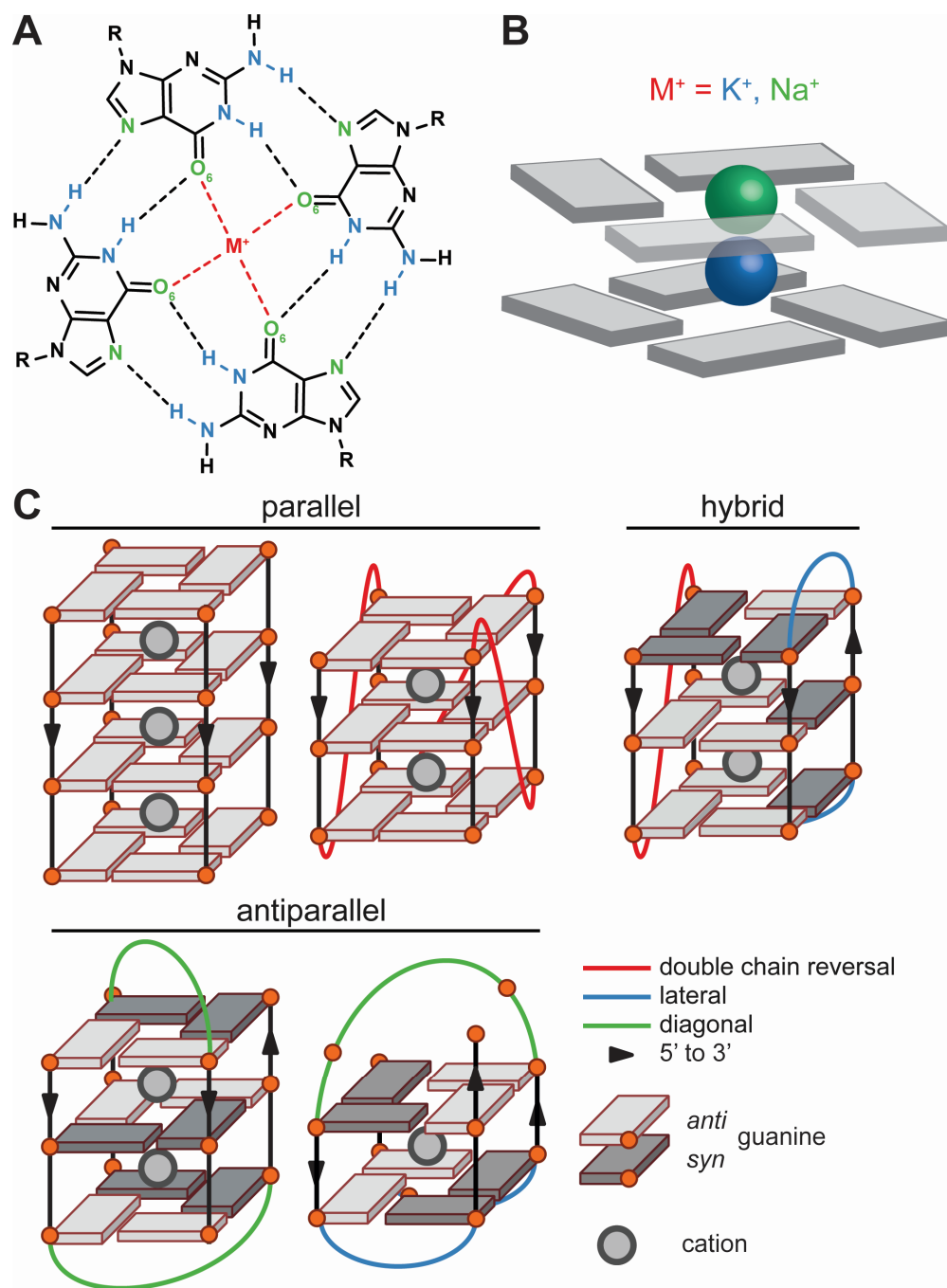
# 1 Introduction: G-Quadruplex nucleic acids

G-quadruplexes (G4) encompass guanine-rich nucleic acids that fold by forming successive quartets of guanines (also called G-tetrads), stabilized by intra-quartet hydrogen bonds (Fig. 1A), inter-quartet stacking, and cation coordination (Fig. 1B). This specific although highly polymorphic type of secondary structure deviates significantly from the classical B-DNA duplex [1-3]. Such G-rich sequences are found in telomeres, and at a statistically remarkable frequency in other part of the genome, notably in promoters [4-6]. G-quadruplexes are detectable in human cells and are strongly suspected to be involved in a number of biological processes at the DNA and RNA levels [7-15]. The vast structural polymorphism exhibited by G-quadruplexes (see section 1.1), together with their putative biological relevance, makes them attractive therapeutic targets compared to canonical duplex DNA [16-20]. A very large – and exponentially increasing – number of studies have consequently been dedicated to G-quadruplexes involved in telomeric sequences, oncogenes, 5'-untranslated regions (5'-UTRs) to cite the most common [21]. Most studies so far involve potassium or sodium cations because of their prevalence in human cells, but a number of other monovalent and divalent cations promote quadruplex formation (*vide infra*). However, among alkali metals, only potassium, sodium and rubidium are truly effective at stabilizing G4s.

This chapter focuses on the essential and specific coordination of alkali metal cations by G-quadruplex nucleic acids, and most notably on studies highlighting cation-dependent dissimilarities. Section 1 surveys G-quadruplex structures and their interactions with alkali metal ions while section 2 presents analytical methods used to study G-quadruplexes. The influence of alkali cations on the stability, structure, and kinetics of formation of G-quadruplex structures of quadruplexes will be discussed in sections 3 and 4. Section 5 focuses on the cation-induced interconversion of G-quadruplex structures. In sections 3 to 5, we will particularly emphasize the comparisons between cations, most often  $K^+$  and  $Na^+$  because of their prevalence in the literature and in cells.

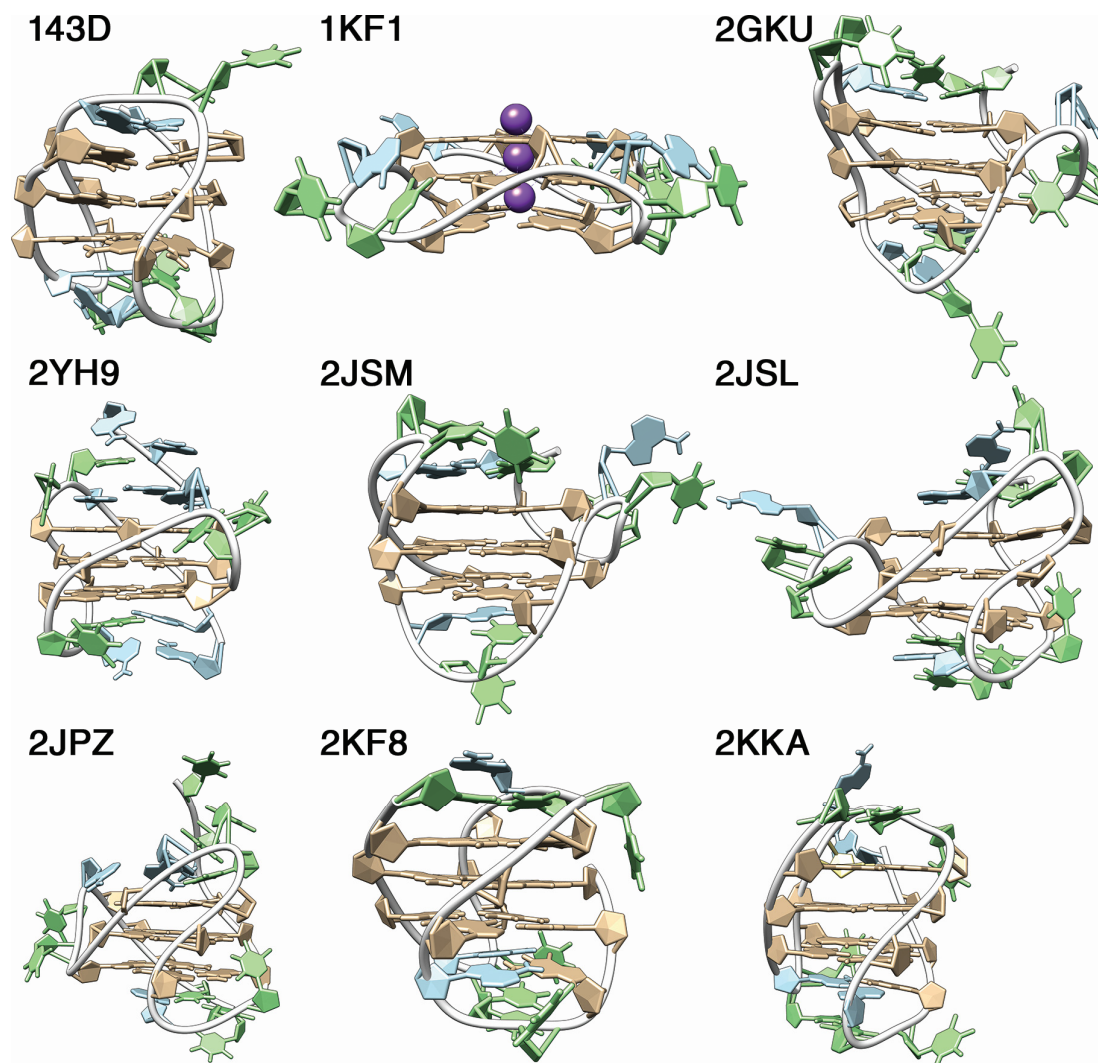
## 1.1 Overview of Structure

The core structure of all G-quadruplexes is constituted by the stacking of at least two G-tetrads (or G-quartets), each formed by the *quasi* co-planar association of four guanines linked by a network of eight hydrogen bonds (Figure 1A,B). The guanines forming the tetrads coordinate some mono- or divalent cations in their center (referred to as central stem), typically  $K^+$  and  $Na^+$ , via their oxygen O6. The single-strand sections linking the guanines tracts are called loops, and can adopt a variety of geometries (Figure 1C): lateral (or edgewise), diagonal, and double chain reversal (or propeller). Furthermore, the four guanine tracts can be oriented (regarding their 5' to 3' polarity) into four different topologies: parallel (the four strands share the same polarity), antiparallel (two strands in a way, the two other in the opposite way; these pairs may correspond to two adjacent or diagonally-opposed strands, resulting in very different geometries), or hybrid (three strands sharing the same polarity, and the last one the opposite) [22,23].



**Figure 1** (A) Guanine tetrads contain four guanines linked by eight hydrogen bonds (donor and acceptor groups in blue and green, respectively). Guanine O6 selectively coordinate a metal cation (red). (B) Tetrads can stack to form G-quadruplexes. Cations of larger ionic radii are located between the tetrads (case of  $K^+$ ; blue), while smaller ones can also coordinate within the plane of the tetrads, or assume an intermediate position (case of  $Na^+$ ; green). (C) G-quadruplexes can fold into a variety of topology, that differ mainly by the relative orientation and number of strands (1 to 4), the number of tetrads (at least 2), and the geometry of the loops.

In antiparallel topologies, the strands sharing the same direction can be diagonally opposed or adjacent. This distinguishes ‘chair’-type G-quadruplexes (three lateral loops, e.g. the thrombin binding aptamer [24]) and ‘basket’-type G-quadruplexes (lateral-diagonal-lateral loops, e.g. the human telomeric sequence d[AG<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>] in Na<sup>+</sup> conditions [25]; Figure 2, 143D) [25-32]. Intramolecular parallel structures (three double-chain reversal loops) are sometimes referred to as ‘propeller’ topologies (e.g., the crystal structure of d[AG<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>] in K<sup>+</sup> conditions [26], Fig. 2, 1KF1).



**Figure 2** Examples of human telomeric G-quadruplex structures deposited in the PDB that were solved by NMR or X-ray crystallography: 143D [25], 1KF1 [26], 2GKU [27], 2YH9 [28], 2JSM and 2JSL [29], 2JPZ [30], 2KF8 [31], and 2KKA [32]. Guanosines are depicted in brown, inosine in yellow, adenosines in blue, thymidines in green, the phosphate backbones as white ribbons, and K<sup>+</sup> as purple spheres where available. All structures except 143D (Na<sup>+</sup>) were solved in K<sup>+</sup> conditions.

The glycosidic bond angle of guanines involved in tetrads can adopt an *anti* or *syn* geometry depending on the relative strand orientation: parallel topologies contain almost exclusively *anti* guanines (for exceptions, see [33-35]), whereas a mixture of *syn* and *anti* is observed for antiparallel and hybrid structures [36]. Consequently, parallel G-quadruplexes generally contain *anti/anti* stacks (*syn/syn* is not favored; see section 1.2), antiparallel ones have *anti/syn* and *syn/anti*, and hybrid ones can have all of the above.

Furthermore, G-quadruplexes may be formed by the folding of a single strand or the association of two to four strands, and some sequences have a tendency to oligomerize [37-39]. One should therefore not confuse the term “quadruplex” or “G-quadruplex” (designating any structure containing stacked guanine tetrads) with the strand molecularity (for which we will adopt the nomenclature “intramolecular”, “bimolecular”, “trimolecular” [40], and “tetramolecular”).

Finally, a number of uncommon features have been observed, such as strand bulges and snapbacks [41-43], base-pairing in loops [44], alternative tetrads [45], and other stacked planar entities (triads, pentads, hexads, heptads, and octads) [31,46-52]. The combination of different loop geometries, strand orientations, and molecularities, implies that G-quadruplexes display an important polymorphism. It has been reported that there is a theoretical number of 26 possible G-quadruplex topologies, not taking into account unexpected folds (*e.g.* isolated guanines involved in tetrads [53]), but only a few of them (six) have been observed *in vitro* [22,23], possibly meaning that some are energetically disfavored, thermodynamically or kinetically. A number of insightful review articles and books cover the G-quadruplex structures published so far [3,12,36,54-59].

## 1.2 Stabilizing Interactions

The factors contributing to the stabilization of G-quadruplexes and yielding a particular folding topology or mixture of topologies are: stacking interactions, hydrogen bonding, solvation, and cation binding. Cation effects on G-quadruplexes differ significantly from those on duplexes (see chapter 6). Cation *coordination* is indeed absolutely required to form G-quadruplexes, to stabilize the G-tetrad stacks. Moreover, sufficient ionic strength is required to compensate electrostatic repulsion between the phosphate oxygens of four strands in G-quadruplexes, instead of two for duplexes (loops may be considered as unfolded single strands depending on the structure).

The total free energy can be decomposed in a number of free energy contributions (*e.g.* Coulombic forces, hydrogen bonding, hydration, van der Waals terms), which can themselves be decomposed in entropic and enthalpic contributions. Analysis of G-quadruplex melting and calorimetric data revealed that G-quadruplex formation is enthalpically driven [60,61]. This results from a more negative (favorable) enthalpy of tetrad formation, only partially compensated by more negative (less favorable) entropies of tetrad formation. Below are presented some key elements to understand G-quadruplex stabilization, and cation coordination *per se* is explored in the section 1.3. An excellent review discussing in details the stability of G-quadruplexes has been published by Lane *et al.* [62].

### 1.2.1. Stacking

Similarly to other nucleic acid secondary structures, G-quadruplex stabilization relies in part on  $\pi$ - $\pi$  stacking of aromatic bases, and more precisely of guanine from consecutive tetrads. Molecular mechanics simulations have predicted the relative stability order among guanine stacks to be (from 5' to 3'): *syn/anti* > *anti/anti* > *anti/syn* > *syn/syn* [63]. This is not fully consistent with the structures solved so far as it does not explain the parallel orientation of the strands in tetramolecular assemblies, possibly because of force-field biases. Quantum mechanical (QM) dispersion-corrected density functional theory (DFT-D3) calculations on stacking of two tetrads, containing one K<sup>+</sup> cation gave a more consistent picture with the structures solved so far: *anti/anti* > *syn/anti* > *anti/syn* > *syn/syn* ( $\Delta E$  vs *anti/anti* = 1.2, 3.5 and 7.8 kcal.mol<sup>-1</sup>, respectively) [64]. It was also suggested in the same study that 5'-terminal H-bonds present a stabilizing effect on 5'-terminal *syn* guanines.

Ultimately, the topology of a G-quadruplex depends on its precise sequence, which can provide additional stabilizing interactions (see below), and on buffer conditions (cations, co-solvents). Because the stacking interactions likely account for a large part of the net energetic gain, increasing the number of quartets is energetically favorable. The number of consecutive tetrads is typically equal to the length of the guanine tracts, although exceptions have been observed. For instance, the 22 mer of the human telomeric sequence d[(G<sub>3</sub>T<sub>2</sub>A)<sub>3</sub>G<sub>3</sub>T] (PBD ID: 2KF8) contains four tracts of three guanines, but folds predominantly in a two-tetrad G-quadruplex (Figure 2) [31].

The formation of higher-order G-quadruplex structures (usually dimers) *via* the stacking of external tetrads of monomer units also provides additional stabilization. Almost all published dimer structures exhibits a 5'-5' interface [65], although 3'-3' stacking remains possible [66,67]. QM and 100-ns MD simulations suggest that while 5'-5' interface readily stack in a favorable manner (60—65° rotation), the 3'-3' interface cannot reach the same type of geometry because of guanine-sugar clashes and therefore yields less favorable stacks (45° or 30° rotations,  $\Delta E$  vs 5'-5' = 4 and 10 kcal.mol<sup>-1</sup>, respectively) [65].

Stacking of other nucleotides (non-tetrad bases) might also contribute to the overall stabilization of the G-quadruplex [31,68]. For instance, in the aforementioned two-tetrad G-quadruplex 2KF8, loop nucleotides provide G•G•G and A•G•G triads, which stack on both tetrads [31]. Alternative planar entities have been identified, such as G•C•G•C tetrads [45], as well as the larger pentads [46,47], hexads [48-50], heptads [51] and octads [52] that provide additional hydrogen bonds. Finally, formation of base pairs is also possible [69]. Searle *et al.* have described a bimolecular G-quadruplex whose two loops form mini-hairpin motifs [44]. The above mentioned 2KF8 structure involves mismatched T•T base pairs [31]. Base-pairing can mediate the formation of stacked G-quadruplex dimers, as observed by NMR for the c-kit2 sequence that exhibits an A•A base pair at the interface [70].



### 1.2.2. Hydrogen bonding

Hydrogen bonding is another stabilization element found in biomacromolecules, and in particular between bases in nucleic acids, typically interacting through two (A•T/U) or three (G•C) H-bonds. In such Watson-Crick base pairing, the stability of H-bonds is linked to donor-acceptor orbital interactions and polarization of the  $\pi$  system by partially neutralizing the charges in the  $\sigma$ -electrons system. In G-quadruplex nucleic acids, there is a network of (C2)NH<sub>2</sub>:N7 and O6:N1H hydrogen bonds. As acceptor and donor groups are likely hydrogen-bonding to water for unfolded/unassociated strands, there is only a small enthalpy change, but the release of water molecules to the bulk medium may induce an entropy gain [62].

A combination of high-resolution variable-temperature STM and DFT calculations suggested that the stabilization of a tetrad induced by the network of eight hydrogen bonds is higher than the sum of four individual G•G pairs [71]. This cooperativity was first ascribed to  $\pi$  assistance, however, DFT-D calculations by Fonseca-Guerra *et al.* suggested that the cooperativity in guanine tetrads more likely arises from the charge separation that is associated to charge transfer between pairs of guanine, *i.e.* donor-acceptor orbital interactions in the  $\sigma$ -electron system [72]. This results in an interruption of the  $\pi$ -electron system, and an enhancement of both the positive charges on the H atoms of H-bond donor groups and negative charges of N and O atoms of acceptor groups. In the same study, it was evidenced that in aqueous solution, the hydrogen bond energy is significantly diminished as compared to the gas phase (-34 vs -80 kcal/mol). Compared to solvation, stacking of three tetrads (in water) has a weak favorable (-2 kcal/mol) effect on the cooperativity. However, introduction of two Na<sup>+</sup> cations in this system entirely restores the stability observed in the gas phase, although it weakens the hydrogens bonds as compared to the system without cation.

## 1.3 Alkali Metal Ion Coordination in G-Quartets

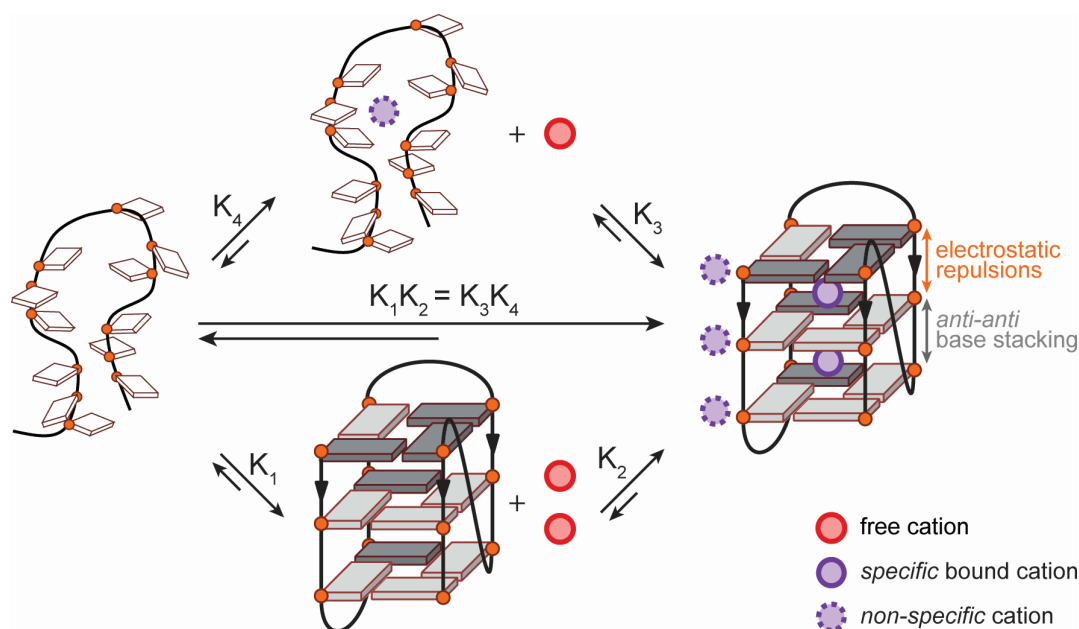
### 1.3.1 Cation preference

Pinnavaia and co-workers' pioneering work on 5'-GMP provided the first indication of the formation of anionic cavities in G-quadruplex-type structures that can selectively complex cations small enough to fit, but large enough to *bridge* the carbonyl oxygens [73], analogously to the binding of metal alkali to crown ethers [74]. This led the authors to propose the complexation of Na<sup>+</sup> in the plane of tetrads, while potassium would fit in the interplanar spacing and be classically octa-coordinated. Follow-up studies [75,76] expanded the scope to four-stranded poly(G) strands [77], helical poly(I) structures including the binding of ammonium [78-80].

A seminal report by Blackburn *et al.* suggested the existence of "G•G base pairs" in the telomeric sequence of several organisms [81]. Following these findings, Williamson,

Raghuraman and Cech proposed a model for telomeric sequences where they formally described the G-quadruplex structure for the first time [82]. The studies conducted on *Oxytricha* and *Tetrahymena* telomeric sequences provided the first hints regarding the critical importance of monovalent cations for G-quadruplex formation, stoichiometry, and stability. It was postulated that  $\text{Li}^+$  is too small to bind in the center of G-quadruplex tetrad, whereas  $\text{Na}^+$  fits perfectly, and larger ions such as  $\text{K}^+$  and  $\text{Cs}^+$  bind in the interplane cavity [82]. It is now widely accepted that cation coordination is essential for the stabilization of G-quadruplexes. By compiling a number of studies, one can estimate that G-quadruplex stabilization follows the general trend:  $\text{Sr}^{2+} > \mathbf{K}^+ > \text{Ca}^{2+} > \text{NH}_4^+, \mathbf{Na}^+, \mathbf{Rb}^+ > \text{Mg}^{2+} > \mathbf{Li}^+ \geq \mathbf{Cs}^+$  (alkali cations in bold) [83-85].

Although the bulkiness of these cations is certainly critical for binding within a G4 cavity, and is historically presented as the main explanation for the cation-dependent stability differences (*vide supra*), other factors play an important role in the stabilization of the complexes [86-89]. Cation binding to G-quadruplexes can be classified as *non-specific* (diffuse [90]), where the cations retaining their outer-sphere hydration bind the negatively charged phosphates, or *specific* (site-bound) by coordination to the guanine O6, where the hydration sphere has been lost and the binding follow the law of mass action (Figure 3). For instance, Gray and Chaires have determined that, at concentration of  $\text{K}^+$  above 2.5 mM,  $\text{d}[\text{AG}_3(\text{T}_2\text{AG}_3)_3]$  attracts up to 6—8 more cations than the predicted two specific binding sites [91]. The binding of these non-specific cations to the negatively charged phosphate backbone reduces the electrostatic repulsions and thus also promotes folding. Note that external coordination to G-quadruplex is not necessarily diffuse: specific external coordination to loops has been suggested for  $\text{Tb}^{3+}$  on the basis of CD and luminescence data [92]. Notably, cations coordinated between the top G-quartet and the loop are often detected in the density maps provided by X-ray crystallography (see section 1.3.3). By comparing  $\text{K}^+$  binding to the 22-mer G-quadruplex  $\text{d}[\text{AG}_3(\text{T}_2\text{AG}_3)_3]$  with the binding to the unfolded  $\text{d}[\text{T}_{22}]$ , it was suggested that the G-quadruplex may specifically bind up to five cations on external sites [91]. However, the use of a control oligonucleotide with such a different sequence and different structure (random coil *vs* globular) may not mimic properly the diffuse binding of the G-quadruplex.



**Figure 3** Multiple equilibria involved in cation coordination and G-quadruplex folding. Each cation can either bind to a preformed G-quadruplex ( $K_1$  then  $K_2$ ), or bind to the random coil that subsequently fold in a G-quadruplex ( $K_4$  then  $K_3$ ). The folded species on the right shows the differences between specific and unspecific binding as well as other stabilizing (stacking) and destabilizing (phosphate repulsions) factors.

### 1.3.2 Cation binding energetics

In specific coordination to the tetrads, the cations are involved in electrostatic and donor-acceptor orbital interactions with the lone pairs of guanine O6, yielding tight  $M^+-O$  coordination bonds [72]. Stabilization is also provided by a screening of electronic repulsion of these O6 lone pairs [93]. In solution, potassium is typically hexa-hydrated, and hence the coordination *per se* by guanine O6 instead of water molecules shall not provide a large enthalpy change. Alkali cations specifically bind within the electronegative cavity formed by the O6 of tetrad guanines in distinct fashions.

As with crown ethers, sodium and potassium bind G-quadruplexes with different affinities, although other factors than the cation radii account for it. The difference in free energies is estimated to be around  $2 \pm 0.5$  kcal.mol<sup>-1</sup> [88,94-97]. The thrombin binding aptamer sequence (TBA; d(G<sub>2</sub>T<sub>2</sub>G<sub>2</sub>TGTG<sub>2</sub>T<sub>2</sub>G<sub>2</sub>)), which contain only two tetrads binds a single cation, with a folding constant  $K_{fold}$  measured by monitoring the folding by spectroscopic means of  $1.3 \times 10^7$  and  $5.5 \times 10^5$  M<sup>-1</sup> at 10 °C for K<sup>+</sup> and Na<sup>+</sup>, respectively [98], or an association constant  $K_a$  of  $2 \times 10^5$  M<sup>-1</sup> for K<sup>+</sup>, determined by mass spectrometry [99]. Specific ion binding has been described thermodynamically by Lane *et al.* by multiple equilibrium involving the folded and unfolded G-quadruplex as shown in Figure 3 [62].

Cation coordination to G-quadruplexes is typically accompanied by G-quadruplex folding that brings further stabilization, which is not experimentally distinguishable except in molecular modeling studies (see section 2.1). Regarding the equilibrium constants defined in

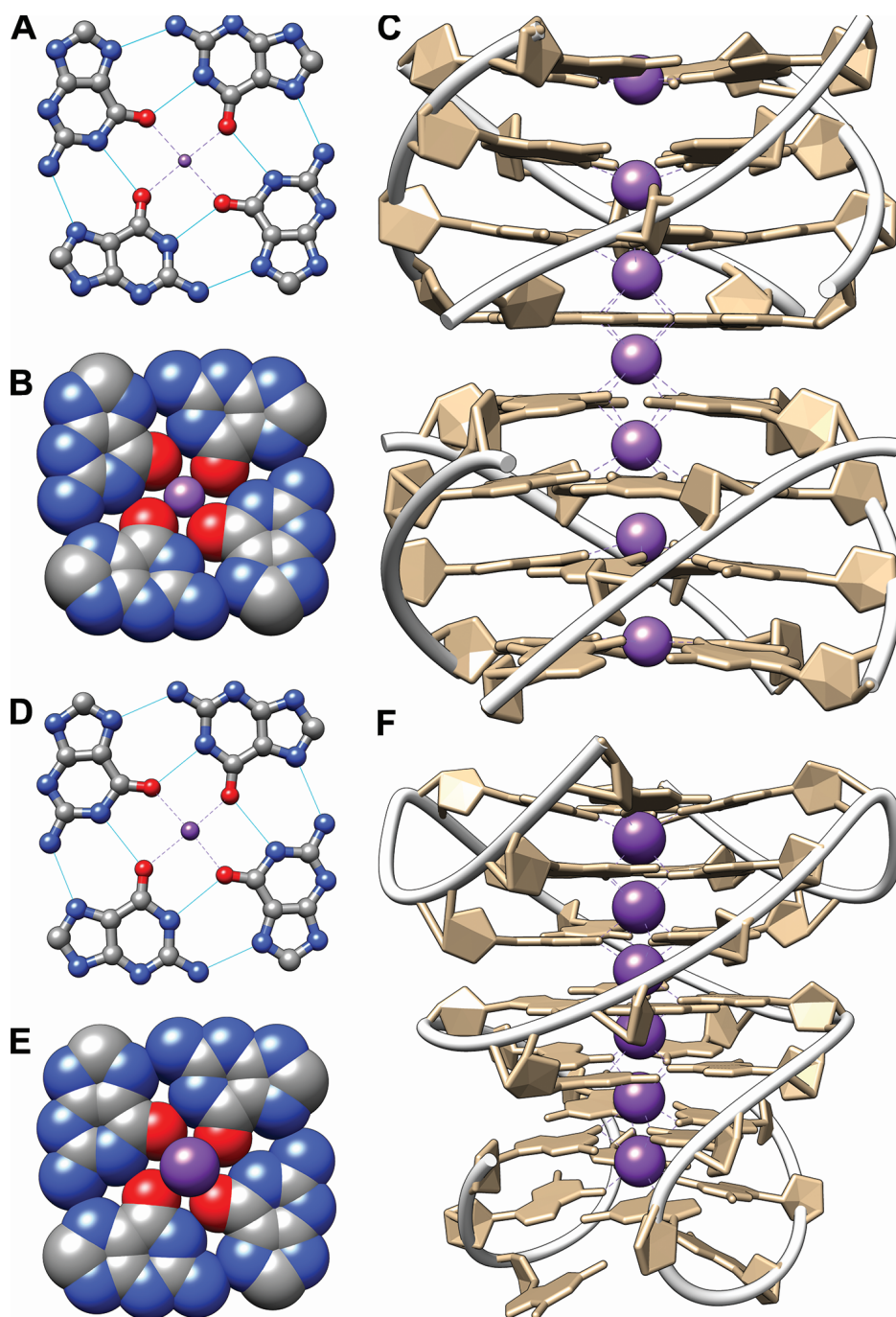
Figure 3, in most cases the  $K_1$  value is large, and the apparent dissociation constant is  $K_1K_2$  (exclusive binding mechanism) [95]. When the solution is devoid of alkali cation, the fraction of folded strand is presumably small. Cation binding to the unfolded strand is also small, and thus an important stabilization is attributed to specific cation binding ( $K_2/K_4$  ratio) [62]. The analogy with alkali cations binding to crown ethers is noteworthy. In aqueous solutions, the crown ether 18-crown-6 binds  $K^+$  and  $Na^+$  with  $\log K \approx 2$  and 0.7, respectively, and complexation enthalpy  $\Delta_r H^\circ$  of -6 and -2 kcal.mol<sup>-1</sup>, at 37 °C, including the desolvation and structural change energies [74,100-102]. Comparatively, the net  $\Delta G$  at 37 °C of three-tetrad G-quadruplex formation has been estimated to 5—10 kcal.mol<sup>-1</sup>, which supports the idea that the energy of potassium binding and of subsequent conformation changes accounts for most of the stabilization [62,91]. For instance, the 22-mer of the human telomeric sequence d[AG<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>] has an overall -2.4 kcal.mol<sup>-1</sup> folding free energy, while in 5 mM potassium,  $K^+$  contributes to roughly -4.9 kcal.mol<sup>-1</sup> [91]. This is also consistent with the unfolded state being largely populated in absence of cation (Figure 3, left). Consequently, binding of potassium or sodium is likely fast and cooperative, whereas the subsequent structural change is quasi irreversible with a net binding energy of around 5 kcal.mol<sup>-1</sup>. This is the case of the human telomeric sequence for instance [95].

The apparent binding of  $Na^+$  and  $K^+$  to human telomeric sequences is very cooperative and lies within 5—15 and 0.5—2 mM, respectively [94,95,103]. Comparatively,  $Na^+$  forms 1:1 complexes with isolated 5'-GMP with a  $K_a$  of 2.85 M<sup>-1</sup> at 5 °C [104]. Feigon *et al.*, and Leszczynski *et al.*, have shown that the free energy of hydration of *specific* cations can explain the stability difference observed between  $Na^+$  and  $K^+$  solutions, the dehydration of the former inducing a greater energetic cost [88,105]. Indeed, unlike diffusely bound cations, tetrad-bound cations lose their whole hydration sphere [106]. The energetics of cation binding to G4s can be decomposed into a positive free energy of dehydration and a negative free energy of coordination *per se*.  $Na^+$  gives favorable coordination energy of coordination, as shown by NMR [88,107], but its binding is penalized by its stronger hydration as compared to  $K^+$ . Meyer *et al.* have also observed by DFT calculations that solvation effects explain the favorable coordination to  $K^+$ , and shown that the cation coordination contribution accounts for 50% of the total interaction energy of a two-tetrad construct [108]. Hydration energy of alkali cations is usually presented as being inversely proportional to their ionic radii.  $K^+$  systematically presents the best compromise, and stabilizes G4s more than  $Na^+$  and  $Rb^+$  (other alkali cations provide very weak to no stabilization). However, the extent to which it does so depends hugely on the studied sequence, and in particular, on the sequence's ability to adopt different structures in presence of different cations (see section 3.2) [85,89,94,109-119].

### 1.3.3 Cation location

It is often stated that  $Na^+$  is small enough (0.95 Å) to fit in the plane of a quartet, while any cation larger than that, such as  $K^+$  (1.33 Å) or  $Rb^+$  (1.52 Å), are coordinated between two planes. In reality, there is a continuum of possible binding sites resulting from (i) the

abovementioned cation lone pair attraction, (ii) the presence of additional cation coordination sites (e.g. loops, dimer interface), and (iii) cation-cation repulsion. Indeed, when more than one cation is bound by the G4, the cations' mutual repulsion also influences their precise locations and the fine structure of the G-quadruplex. Phillips *et al.* showed crystals of  $[d(TG_4T)]_4$  coordinated to  $Na^+$  and  $Ca^{2+}$  that provides a nice illustration of this phenomenon for  $Na^+$  [120] (Figure 4A,B,C) [120,121]. The G-quadruplex is arranged in a head-to-head dimer with eight consecutive tetrads. Out of the seven bound sodium cations, only the two outer ones lie in the G-tetrad's plane, in a four-coordinate fashion, completed by the binding of water molecules. The 5 other sodium cations are equally spaced between the two other ones, and adopt intermediate coordination positions in a continuum ending at the central cation that is sandwiched exactly halfway between two tetrads, with a bipyramidal geometry similar to the one usually adopted by  $K^+$ . As a result, the average distance between  $Na^+$  cations (4.2 Å for outer cations, 3.6 Å for internal cations) is higher than the distance between G-tetrads. A slight distortion of the external quartets is also observed in this crystal, putatively allowing the inter-cation distance to increase. One should note however, that crystal packing forces might account for the position of the cations, which therefore could not reflect the reality of solution-based G-quadruplexes.



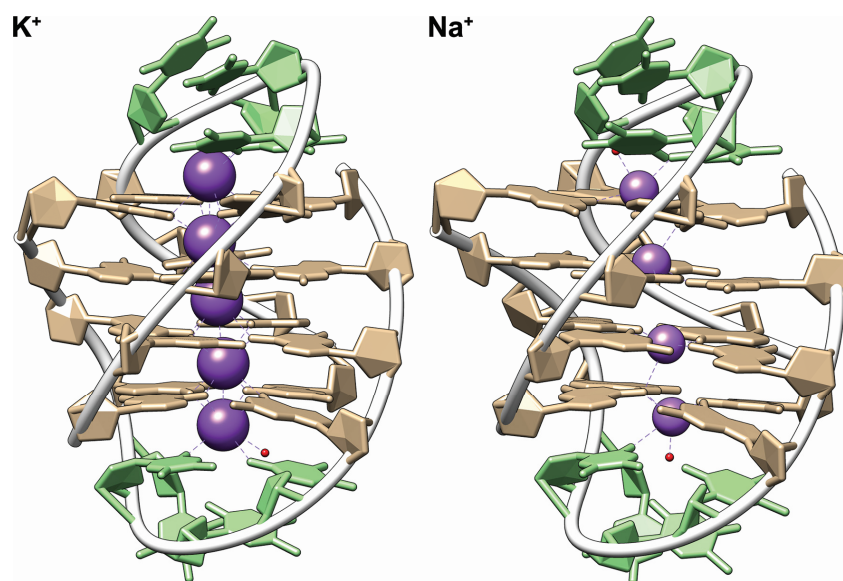
**Figure 4** Crystal structure of  $[d(TG_4T)]_4$  coordinating  $Na^+$  (PDB ID 352D, [120]) (A, B, C), and  $d(G_3CG_4AG_5A_2G_3A)$  coordinating  $K^+$  (PDB ID 4H29, [121]) (D, E, F): tetrad geometry depicted as balls and sticks (A, D) and spheres (B, E), the head-to-head dimer coordinating seven  $Na^+$  adopting positions ranging from within the tetrads (external cations) to sandwiched midway between two quartets (central cation) (C), and the intertwined dimer of B-raf coordinating six sandwiched  $K^+$  cations (F). Hydrogen bonds are shown in blue,  $Na^+$  and  $K^+$  cations and coordination in purple. Carbon atoms are in grey, nitrogen in blue, and oxygen in red except in panels C and F where guanines are colored in brown and the phosphate backbone is a white ribbon. For the sake of clarity, dT, dA, and dC residues,  $Ca^{2+}$ ,  $H_2O$ , and hydrogen atoms are omitted.

Creze *et al.* obtained a similar dimeric structure with Na<sup>+</sup> being coordinated increasingly more within the tetrad planes when going towards the extremities of the G-quadruplex ( $d_{\text{Na}^+-\text{Na}^+} = 3.4\text{--}4.7\text{ \AA}$ ), with water molecules capping at the external tetrads ( $d_{\text{Na}^+-\text{H}_2\text{O}} = 2.4\text{--}2.5\text{ \AA}$ ) [122]. Crystallization was carried out from a sodium solution of quadruplex using lithium sulfate as a precipitating agent. A single Li<sup>+</sup> ion was observed in a groove, and none of them in the central stem.

Incidentally, in the first reported crystal structure of a small molecule (daunomycin) bound to a G-quadruplex, the Na<sup>+</sup> cations are all coordinated in the plane of the tetrads [123]. The tetramolecular [d(TG<sub>4</sub>T)]<sub>4</sub> G-quadruplex is also observed as a dimer, with two daunomycin molecules at the interface between the individual tetramers, which might also be linked to the difference of sodium positioning. In a more recent report on the high-resolution crystal of the dimeric [d(TG<sub>4</sub>T)]<sub>4</sub>/daunomycin complex employing *syn* glycosyl linkages, Na<sup>+</sup> cations are not in the plane of the tetrads, but rather sandwiched in various fashions [124]. Notably, the Na<sup>+</sup> closer to the 5'-interface lies almost midway between two tetrads, with a square prism geometry, due to the *syn* glycosyl orientation of the first guanine residue. Interestingly, Na<sup>+</sup> also occupies a site at a daunomycin-daunomycin interface.

Conversely, in the crystal structure of the intertwined dimeric quadruplex formed by the B-raf sequence d(G<sub>3</sub>CG<sub>4</sub>AG<sub>5</sub>A<sub>2</sub>G<sub>3</sub>A) (PDB ID: 4H29), the six potassium cations are all observed in-between the tetrads, in a quasi linear arrangement, and are equidistant from each other ( $d_{\text{K}^+-\text{K}^+} = 3.44 \pm 0.09\text{ \AA}$ ) (Figure 4D,E,F) [121].

The differences in location of sodium and potassium cations has been extensively investigated by NMR and X-Ray crystallography for the telomeric sequence from *Oxytricha Nova* d(G<sub>4</sub>T<sub>4</sub>G<sub>4</sub>), forming an antiparallel bimolecular G-quadruplexes, where the thymines form two diagonal loops (Figure 5) [96,107,125-128]. The crystal structures 1JPQ and 1JRN exhibits five equidistant K<sup>+</sup> cations, spaced by 3.4 Å on average, three being sandwiched between the tetrads and the two other coordinated between the external quartets and within the loops [128]. The K<sup>+</sup>-K<sup>+</sup> distance is the same that has been measured for the human telomeric sequence [26]. Similar results were observed with Tl<sup>+</sup> cations, which have a similar radius than K<sup>+</sup> (5 bound cations with a 3.6 Å spacing; average RMSD = 0.26 Å) [129]. Conversely, only four Na<sup>+</sup> cations are coordinated, the binding sites in the loops being empty (PDB ID 1JB7) [127]. Sodium cations are coordinated in the planes of the central tetrads while the external cations are bound slightly outer of the tetrads, towards the loops, and are coordinated by the thymines O2. This partial loop coordination leads to a distribution of Na<sup>+</sup> location that differs from [d(TG<sub>4</sub>T)]<sub>4</sub>, but here too the distance between sodium cations is higher than the distance between tetrads.



**Figure 5** Crystal structure of  $[d(T_4G_4T_4)]_2$  coordinating  $K^+$  (PDB ID 1JPQ) or  $Na^+$  (PDB ID 1JB7). Guanines are shown in brown, thymines in green, the cations as purple spheres, and the oxygens from water molecules completing the spheres of coordination as red balls.

Overall,  $K^+$  cations are specifically coordinated in between tetrads, in a nearly octahedral way, or alternatively within loops above tetrads, satisfying the usual hexacoordinate stereochemistry of these cations. Positioning of bases outside of the plane of the tetrad may be necessary to comply with this geometry, to an extent balanced by stacking and hydrogen bonding energies. Smaller  $Na^+$  cations can be coordinated within the plane of tetrads, and can occupy a range of positions owing to lower steric constraints, hence reducing the electrostatic repulsions.  $K^+$  is bound with higher affinity than  $Na^+$  and  $Rb^+$ , while  $Li^+$  and  $Cs^+$  are poorly coordinated by G-quadruplexes. Although the cationic radius certainly accounts for these differences, the energy of hydration has been demonstrated to be the cause of the energetic preference for  $K^+$  vs  $Na^+$ .

All these examples highlight that the stability of G-quadruplexes is massively driven by cation binding and the resulting structural (re-)organization, rather than by other weak interactions.

## 2 Methods to study G-quadruplex nucleic acids

### 2.1 Folding Topology

A number of high- and low-resolution analytical methods are used to determine the topology of G-quadruplexes [130], and some are presented hereafter. NMR spectroscopy, X-ray crystallography, and molecular modeling give access to atomic-scale structure information



[36,131-134], and have revealed over the last decades an impressive structural polymorphism among G-quadruplexes (Figure 2). A number of other spectroscopic and spectrometric methods (e.g. UV absorption spectroscopy, electronic circular dichroism spectroscopy, native mass spectrometry, electrophoresis, and chromatography) are employed to determine strand orientations and stoichiometry, and molecular sizes. As each of these techniques give different types of information, and furthermore presents certain drawbacks, it is advisable to combine a number of these techniques before drawing final conclusions.

### 2.1.1 X-Ray crystallography

X-Ray crystallography gives access to atomic-scale resolution structure of DNA, RNA, or LNA G-quadruplexes, including the cations (see section 2.2.1), water molecules, and binders [132,133]. A fair number of structures containing alkali cations (or Tl<sup>+</sup> as a K<sup>+</sup> surrogate; *vide infra*) have been solved using this method [26,67,120-124,127-129,135-161], mainly by the teams of Parkinson and Neidle, and the late prof. Sundaralingam. Screening of sequences, as well as base modification (e.g. heavy-atom addition) and loop/flanking sequence changes are typically performed in order to obtain a sequence that crystallizes and diffracts well. These aspects, as well as crystallization protocols are discussed in details in reference [133].

Almost all of the available crystal structures are tetrameric [120,122,123,135-145] or dimeric, *via* association of distinct strands or external stacking of monomer units, often templated by binders [26,67,121,124,127-129,145-160]. Moreover, stacked monomer units and intermolecular species [26,158,159] are systematically parallel-stranded within each subunit, except for the thrombin binding aptamer (TBA) bound to  $\alpha$ -thrombin [161]. This raises questions as to the possibility to crystallize structures such as intramolecular antiparallel-stranded or hybrid structures (either because they do not crystallize or because the crystal packing forces induce conversion to parallel and multimeric folds), and this in turns casts some doubts about the complete relevance of this method for solution-based folding studies. Crystallography is nevertheless very powerful when it comes to characterize cation coordination (see section 2.2.1), and more generally ligand binding (reviewed in [162]). It also provides better defined structures than NMR in cases where the latter is limited by internal dynamics or the presence of mixtures of conformations, with or without inter-conversion between structures.

### 2.1.2 Nuclear magnetic resonance (NMR) spectroscopy

High-field NMR is broadly used to study G-quadruplex structures and their dynamics [36,131,163,164]. The full topology of many structures has now been solved, including the human telomeric DNA [25,27-32,45,165-172] and RNA sequences (TERRA) [50,173,174], telomeres from other species [46,68,83,96,125,126,175-177], human oncogenes [53,70,178-182], minisatellites [43,183], and aptamers against biological targets [24,41,184-186] or

viruses [15,69]. Besides, even when solving the full topology is not possible or required, several types of NMR experiments can provide information on particular aspects of the structure.

First, the formation of G-quadruplexes can be affirmed by the presence of guanine imino protons H1 in the 10—12 ppm range, typical of G-tetrads, whereas canonical Watson-Crick base-pairing shifts the protons to 13—14 ppm [134,187]. Moreover, compared to Watson-Crick protons, tetrad protons – particularly the ones from central tetrads – exchange more slowly with water, which results in sharp peaks [125,188].

Second, the number of imino proton peaks is linked to the number of tetrads (one peak per guanine, hence four peaks per tetrad, unless some protons are equivalent) [183], and a number of peaks higher than the number of guanines indicates a mixture of species. The kinetics of quadruplex formation are easily accessible given that the time-scale is compatible with NMR experiments [164].

To fully solve a structure however, a number of issues must be tackled: formation of multiple species by a single sequence, higher-order structures, ambiguous assignments [131]. Sequences folding in multiple conformers (e.g. [167,170]) can be characterized if the peaks are sufficiently resolved, or more frequently by screening for or favoring a given conformation by sequence modification (flanking and loop bases, modified bases [170,172]), sample preparation, and buffer and cationic conditions (reviewed in [131]). Stoichiometry can be studied by titration [170,176], or diffusion ordered spectroscopy (DOSY) experiments [189,190], or alternatively by other methods described hereafter (ESI-MS, SE-HPLC, PAGE). Peak assignments was historically performed by through-space NOE experiments [125], possibly helped by base modifications (typically, guanine to inosine or bromoguanine, or thymine to uracil). Site specific  $^{15}\text{N}$ - and  $^{13}\text{C}$ -enrichment of defined residues allow the unambiguous elucidation of G-quadruplex structures routinely [191]. Fold determination can be typically performed by studying either the connectivity between guanine H1 and H8 from the following residue in NOE patterns, the J-couplings through tetrad hydrogen bonds [131]. Hydrogen bonds have been characterized by the H-bond scalar coupling [107].

### 2.1.3 Molecular modeling

Molecular modeling is a third way to get atomic-scale information on the structure of quadruplexes. Advantageously, it gives access to individual energetic terms, dynamics, intermediates, and does not suffer from mixtures of conformations because it studies one molecule at a time [64,65,192-200]. As a result, the structure, energetics and dynamics of rare conformations or of putative reaction intermediates, which would not be possible to isolate experimentally, can be explored. The quality of force fields is however an issue, notably the treatment of the loops, inter-cation repulsion, and the treatment of specific and non-specific electrostatic interactions [193-196,201,202], and combination with biophysical data can prove useful [203].

#### 2.1.4 Electronic circular dichroism (CD) spectroscopy

The most common and straightforward low-resolution technique for the study of G-quadruplex topology is circular dichroism (CD) [204-207]. CD is a spectroscopic method that measures the difference in absorption of left- and right- circularly polarized light by chiral compounds or by compounds in a chiral environment. G-quadruplex nucleic acids are typically characterized by their bands in the UV area (210—300 nm), arising from electronic transitions between stacked guanines. More precisely,  $\pi$ -  $\pi^*$  transitions within guanine exciton couplets occur at 279 and 248 nm, giving rise to CD bands. The sign and intensity of these bands is dictated by the relative orientation of the stacked guanines (head-to-head: *anti/syn*, *syn/anti*; head-to-tail: *anti/anti*, *syn/syn*).

As a result, the G-quadruplex topology can be inferred from the CD spectrum by analysis of the characteristic bands. Parallel topologies exhibit an intense positive bands centered at 260 nm and a negative bands at 240 nm, antiparallel topologies are characterized by a positive band at 290 nm, a (sometimes shallow) negative band at 260 nm, and a positive band at 245 nm, and hybrid-type structures have a positive band at 290 nm, and a shoulder at 260—270 nm. CD is relatively fast and easy to perform but has a low resolution as it only provides a global information on guanine stacks from the sample. It is not possible to easily distinguish mixtures of structures from pure species. For instance, mixtures of antiparallel and hybrid or parallel structures might lead to spectra that can be confused with a pure hybrid-type signature (see the controversy over the topology of d[AG<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>] in section 4.1.1).

Finally, other supramolecular features (additional base stacking, double helices, mismatches) can account for the overall spectrum and possibly lead to misinterpretations [205]. An example of additional guanine stacking modifying the CD signature is provided by the 2KF8 structure, described in section 1.2 (Figure 2) [31]. Although this G-quadruplex has a 2-tetrad antiparallel topology where the tetrad guanines are stacked in a *syn/anti* fashion, its CD signature exhibits a positive band at 260 nm that could be mistakenly interpreted has a hybrid-type topology, due to additional base stacking on both tetrads of triads composed of loop nucleotides, yielding *anti/anti* stacks [205]. Additionally, non-classical tetrads (such as G•C•G•C as in the human mutant telomeric GGGCTA motif) may also alter the CD spectra [45]. Similarly, the presence of two loops forming a mini-hairpin motifs in an antiparallel G-quadruplex leads to the presence of a wide peak centered at 280 nm, characteristic of a B-DNA duplex, masking the expected minimum at 260 nm [37,44].

#### 2.1.5 UV absorption spectroscopy

The UV absorption spectrum slightly differ between single structures, and differences in topologies are usually appreciated by *variations* in absorbance spectra, monitored by difference spectra between folded and unfolded forms. For example, thermal difference spectra (TDS) fingerprints monitor the difference between unfolded spectrum (high

temperature) and the folded spectrum (low temperature) [205]. TDS fingerprints of G-quadruplexes typically exhibit local maxima at 240, 255, and 275 nm, and a minimum at 295 nm [113]. As with CD, different guanine-guanine stacks may indeed modify the intensity and wavelength of absorption bands. Hence, parallel topologies appear to be discriminated from hybrid and antiparallel topologies thanks to their greater  $\Delta A_{240\text{nm}}/\Delta A_{295\text{nm}}$  ratio.

The use of TDS signatures is however not optimal because high melting temperatures leading to decreased TDS intensities, and temperature-dependent changes in molar absorption coefficient both contaminate the signature. Isothermal difference spectra (IDS), which are calculated by subtraction of the UV-spectra of a given sample acquired in absence or presence of cation, at constant temperature. IDS are not strictly identical to TDS because they do not suffer from the drawbacks mentioned above. It is possible to study kinetics of folding and interconversion by this method [208]. In both cases the main advantages are the high throughput and ease of use that compensate their low resolution. Coordination, can be indirectly monitored by either of these methods because it translates into quadruplex folding. Displacement of a cation by another can be observed by IDS, although the signature can differ significantly from case to case, and a significant change in guanine stacking is required for a signal to be observed [208].

UV-melting at 295 nm is a very common methodology to assess the stability of G-quadruplexes [209,210]. As mentioned above, the formation of G-quadruplexes lead to a change in absorbance at 295 nm that can be followed as a function of temperature: increasing temperatures induce an unfolding of the G-quadruplexes that translates into a hypochromism at 295 nm. Alternatively, oligonucleotides tagged with two fluorophores compatible with Förster resonance energy transfer can be used for melting experiments (FRET-melting) [94,98,211,212]. This is not the favored technique for structural studies since the fluorophores tend to form stacked exciton couplets that might affect both the structure and the stability of the G-quadruplex. Whatever the method used to detect the fraction folded, cation binding is related only indirectly to the fraction folded, monitored by the melting curve. Because the relationship is indirect, thermodynamics parameters extracted from such curves do not directly give access to the  $K_d$  of coordination.

### 2.1.6 Separative techniques

Native gel electrophoresis can be employed to determine strand stoichiometry, although it suffers from a number of drawbacks (charge screening effects, smearing, and tedious/ambiguous quantification). It was recently combined to CD for the study of multimers [38]. Size-exclusion HPLC (SE-HPLC) was shown to be a valuable alternative to get insight into oligonucleotides secondary structures, and most notably the strand stoichiometry of quadruplexes [37,164,213-215]. The salient advantages of this technique is the possibility to work in a variety of conditions including cation nature and concentration, and various strand concentrations ( $\mu\text{M}$ — $\text{mM}$ ) that are not all accessible by NMR and X-Ray crystallography. Additionally, it is possible to work with mixture of structures, which are abundant with G-

quadruplex-forming sequences, whereas it is more difficult with techniques such as NMR or CD. Thermodynamics and kinetics of G-quadruplex folding can be studied with this method. It however does not provide information regarding the strand orientation or the number of quartets. Analytical ultracentrifugation (AUC) is a thermodynamically rigorous approach for the determination of absolute molecular weights and get insight into the hydrodynamic shapes of macromolecules in solution [216], and is thus broadly used for the determination of topology and stoichiometry of quadruplexes [203,214,217-223].

### **2.1.7 Native mass spectrometry**

Mass spectrometry (MS) in non-denaturing conditions (both in solution and in the mass spectrometer) is called native MS. Electrospray ionization (ESI) ensures non-denaturing ionization and vaporization. Native ESI-MS is another straightforward way to determine strand stoichiometry [39,224], as well as kinetics and thermodynamics of folding of quadruplexes [225], from pure species or from mixtures, without however getting direct information on the structures. Matrix-assisted laser desorption/ionization (MALDI) has also been used to determine the strand stoichiometry [173], but this ionization method has higher risk of denaturing the complexes than ESI [226,227].

Interestingly, the number of potassium cation bound by a given G-quadruplex can be determined by ESI-MS (see section 2.2.3 for more details), which can give insight into its folding topology [228,229]. Indeed, given the preferred location of specifically bound  $K^+$  cations, the number of tetrads of a G-quadruplex is usually equal to the number of coordinated  $K^+$  cations plus one. In order to get additional insights into structures of macromolecules, ion-mobility spectrometry (IMS) can be coupled in-line to ESI-MS [230]. ESI-IMS-MS is a technique that, in addition to mass-to-charge ratio, further separates ions on the basis of their mobility in a drift tube filled with gas (typically, nitrogen or helium), where a static electric field is applied [231]. The travel time is related to the collision cross section of the ions with the gas: compact structures collide less with gas and travel faster as a result. Gas phase structural determination by MS- and spectroscopy-based methods have been reviewed in a recent book [232].

### **2.1.8 Miscellaneous**

Other methods have been employed to study the topology, stability, and formation of G-quadruplexes, as well as cation binding, albeit tangentially, including single molecule FRET [233], fluorescence melting [94,234-236], denaturing PAGE [237,238],  $^{125}I$ -radioprobeing [239,240], electron paramagnetic resonance (EPR) [222], surface plasmon resonance [241], fluorescence spectroscopy [242,243], Raman spectroscopy [244,245], calorimetry (DSC, ITC) [61,84,112,246-248], temperature-jump relaxation experiments [62,170,236,249,250]. G-quadruplex characterization assays are reviewed in references [62,130,251,252].

## 2.2 Cation Coordination

### 2.2.1 X-Ray crystallography

X-ray crystallography is an invaluable method for the study of cation binding in G4 nucleic acids because it directly locates cation binding sites through electron density. Crystal packing forces may however change the location of the cation as compared to solution-state structures [107,190].  $\text{Tl}^+$  cations can be used as probes for potassium binding sites.  $\text{Tl}^+$  and  $\text{K}^+$  have close atomic radii (1.44 and 1.33 Å, respectively), dehydration energies and bond lengths, and can therefore lead to similar folds, yet  $\text{Tl}^+$  exhibits a higher X-ray scattering potential [142]. Crystals obtained in the presence of  $\text{K}^+$  cations can be converted to  $\text{Tl}^+$ -crystals by soaking in thallium acetate [129].

### 2.2.2 NMR spectroscopy

NMR provides evidence not only on the location of cations within G-quadruplex structures but also on their dynamics [97,163]. First examples of alkali cation coordination by tetrads of 5'-GMP, detected by NMR involved  $^{23}\text{Na}^+$ ,  $^{39}\text{K}^+$  and  $^{87}\text{Rb}^+$  [75,253,254]. Notably, the resonance of  $^{23}\text{Na}^+$  cations undergoes a significant line broadening and up-shifting upon binding [254]. The detection of these cations was extended to G-quadruplexes by Wu *et al.* [190,255]. The coordination and dynamics of sodium cation coordination can also be directly monitored by solution- and solid-phase  $^{23}\text{Na}$  NMR experiments [190,256,257]. In the same vein,  $^{39}\text{K}$  exhibits a characteristic signature on solid-state magic-angle spinning (MAS) NMR spectra upon coordination by G-quadruplexes [258].

Indirect observation of potassium binding has been performed based on proton NMR, chemical exchange and dynamic analysis on d(G<sub>3</sub>T<sub>4</sub>G<sub>3</sub>) and the *Oxytricha* telomeric sequence d(G<sub>4</sub>T<sub>4</sub>G<sub>4</sub>) bound to  $\text{Na}^+$  and  $\text{K}^+$  [88,96]. Another indirect mean of studying  $\text{K}^+$  coordination is the use of  $\text{Tl}^+$  cations as surrogate, in a strategy reminiscent of the one used in X-Ray crystallography studies (*vide supra*).  $^1\text{H}$ - $^{205}\text{Tl}$  scalar couplings provide important constraints for structure determination [129,259].

$^{15}\text{NH}_4^+$  dynamics can be monitored directly with heteronuclear correlations exchange spectroscopy ( $^{15}\text{N}$ - $^1\text{H}$  N<sub>z</sub>ExHSQC) [97], and serve as a reporter probe. Ammonium can indeed displace  $\text{Na}^+$ , or be displaced by  $\text{K}^+$ , allowing an indirect measurement of  $\text{Na}^+$  and  $\text{K}^+$  movements within G-quadruplexes [260]. Changes in topology by displacement of  $\text{Na}^+$  by  $\text{K}^+$  can also be used to study cation binding, albeit more indirectly [88]. Finally,  $\text{Na}^+$  cation dynamics have been studied by nuclear magnetic relaxation dispersion study [261].

### 2.2.3 Native mass spectrometry

Mass spectrometry seems to be an obvious choice for the study of cation binding to G-quadruplexes, since each coordinated cation increases the mass of the complex. A number of mass spectrometry studies were devoted to  $\text{NH}_4^+$  binding determination [52,224,225,231,262-266], because ammonium acetate is a traditional volatile buffer used in biomolecule mass spectrometry. The study of alkali cations is however more tricky because (i) they tend to form clusters even at submillimolar concentrations, and (ii) they are not volatile like the classically employed ammonium acetate. The former issue leads to very noisy spectra from which no information can be inferred, while the latter means that diffuse (unspecific) coordination of alkali cation is at least partially preserved in the gas phase, which results in a higher number of bound cations than tetrad coordination sites.

These issues have recently been partially tackled by Gabelica *et al.* with the use of trimethyl ammonium acetate (TMAA) as a volatile co-solvent [228]. TMA is too bulky to be coordinated by the O6 of tetrad guanines and is therefore a cation of choice to fix the ionic strength and neutralize the backbone phosphates while allowing to study the specific binding of potassium (or other cations) to the tetrads. Mathematical subtraction of diffusively bound cation can be performed in a second step, by using a reference sequence that does not form a G-quadruplex [229], yielding cleaned mass spectra. Titration of G4-forming sequences with increasing potassium concentrations, and subsequent cleaning of the non-specific adducts, give access to the  $K_d$  of coordination of potassium for each tetrad binding site. This methodology is only applicable to G-quadruplexes stable at low (submillimolar) cation concentrations, which is usually not the case when working with sodium. Gross *et al.* have also developed a method that utilizes the gas-phase signal fractions from the bound and unbound species as input into a mathematical model that determines the binding constants [99]. This method was applied to the measure of the  $K_d$  of potassium (and strontium) to TBA.

Another bulky ammonium derivative, namely tetrabutylammonium phosphate, had already been used for the same purpose by Gray and Chaires in spectroscopic experiments involving the human telomeric sequence [91]. This allows the determination of apparent binding constants by CD. Another salient point of this study is the use of a potassium-binding benzofuran-isophthalate crown ether indicator (PBFI) to determine the concentration of free potassium by fluorescent titration, and hence deduce the concentration of coordinated cations and thermodynamic parameters [91,267].

## 3 Role of alkali metal ions in G-quadruplex stability

The stabilization of G-quadruplexes by cations has been studied for the past fifty years on increasingly complex systems. Decades after the first observation of guanosine gels [268], Gellert and co-workers have reported in 1962 the formation of gels by 5'-GMP in presence of sodium, at acidic pH, and postulated the structuration of guanines in tetrads (or quartets), stacked upon each other to form helices [269]. Similar structures were observed a decade later

using various guanosine derivatives, in presence of potassium [270]. However, the seminal paper regarding the necessary and selective complexation of cations by such structures, at neutral pH, dates from the late seventies [73]. Pinnavaia and co-workers have demonstrated by NMR experiments that  $K^+$ ,  $Na^+$ , and  $Rb^+$  all lead to the formation of stacks of three tetrads of 5'-GMP, in a head-to-tail or alternating head to tail and head to head stacking, whereas  $Li^+$  and  $Cs^+$  were found to be ineffective. Additionally, it was evidenced that potassium induces a higher thermal stability than sodium. In the same manuscript is reported the first qualitative ranking of alkali cation-induced stability for 5'-GMP,  $K^+ > Na^+, Rb^+ \gg Li^+, Cs^+$ , which remarkably contrast with the ranking observed for duplex DNA where the binding constant decreases with increasing ionic radii. Research on that topic moved to biologically relevant sequences, notably the telomeres of various species, and model G-quadruplexes such as  $[d(TG_4T)]_4$ . Hereafter, we will review studies highlighting the influence of alkali cations on the stability of simple tetramolecular G-quadruplexes (section 3.1), and sets of telomeric or artificial sequences (section 3.2). The contribution of folding and unfolding rates in the cation-dependent stability of G-quadruplexes is discussed in the section 3.3.

### 3.1 Case Study: $dTG_{3-5}T$ Tetramolecular G-Quadruplexes

Parallel-stranded tetramolecular G-quadruplexes formed by  $dTG_nT$  ( $n = 3-5$ ) sequences have been used extensively as models. These short sequences form unambiguous G-quadruplex structures devoid of loops that makes them “minimal” G-quadruplexes ideal to conduct basic research. A typical measure of the stability of G-quadruplexes is the melting temperature ( $T_m$ ), usually provided by UV-melting [209,210,271], CD-melting, and DSC [61,84,112,246] experiments. Mergny *et al.* have reported the melting temperatures obtained by UV-melting for a number of DNA and RNA tetramolecular G-quadruplexes [117]. Typically, tetramolecular RNA G-quadruplexes are more stable than their DNA counterparts, thanks to higher association rates and lower dissociation rates. Noteworthy, RNA tetramolecular G-quadruplexes were shown as early as 1991 to be highly stable in  $K^+$ , and more than in  $Na^+$  and  $Li^+$  [272]. For instance, the difference in  $T_m$  is large (35 °C) between  $[r(UG_4U)]_4$  and  $[d(TG_4T)]_4$ , in 110 mM  $Na^+$  solutions (Table 1) [112,117,273,274]. Finally, no dissociation is observed in most cases for  $G_5$  and longer tracts in  $Na^+$ , and  $G_4$  and longer tracts in  $K^+$ , further highlighting the cation effect on the stability of tetramolecular G-quadruplexes.

The short tetramolecular G-quadruplex  $[d(TG_3T)]_4$  is characterized by an apparent melting temperature of 48 °C in presence of 110 mM KCl and 16 °C ( $\Delta T_m = 32$  °C) in presence of 110 mM  $Na^+$  [117]. In the case of slow association/dissociation kinetics, the melting curve is not an equilibrium curve, and therefore depends on the temperature gradient. Melting of intermolecular quadruplexes is also affected by the strand concentration. A close result was obtained in 100 mM  $K^+$  by Balasubramanian *et al.* (51 °C; 200  $\mu$ M strand concentration) [273]. Almost identical stability differences are observed with the addition of flanking bases mimicking the human telomeric sequence, as in  $[d(T_2AG_3)]_4$  (50 vs. 17 °C;  $\Delta T_m = 33$  °C) and  $[d(T_2AG_3T)]_4$  (55 vs. 24 °C,  $\Delta T_m = 31$  °C) [117]. NMR experiments on a



high d(T<sub>2</sub>AG<sub>3</sub>) strand concentration (6.4 mM), in a 110 mM KCl buffer, NMR produced a  $T_m$  of around 60 °C [274].

The stability of the longer [d(TG<sub>4</sub>T)]<sub>4</sub> has been studied by different groups [89,112,117]. In presence of 110 mM KCl, the melting temperature is higher than 94 °C [89,117], indicating a very high thermal stability. An increase in Na<sup>+</sup> concentration did not affect significantly the stability ( $\Delta T_m < 2$  °C in the 50—400 mM Na<sup>+</sup> range), but changed dramatically the association rate (see section 3.3) [117]. A large difference ( $T_m = 75$  °C) can be noticed with another study by Petraccone *et al.* [112], performed in 210 mM NaCl but at a higher strand concentration (160  $\mu$ M). Addition of thymine at both termini [d(T<sub>2</sub>G<sub>4</sub>T<sub>2</sub>)]<sub>4</sub>, substitution of the 5'-dT by a dA [d(AG<sub>4</sub>T)]<sub>4</sub>, or use of the RNA counterpart [r(UG<sub>4</sub>U)]<sub>4</sub>, all increase the thermal stability in 110 mM Na<sup>+</sup> ( $T_m = 71, 60,$  and  $89$  °C, respectively) [117]. Further increase of the guanine tract leads to even more stable G-quadruplexes: the melting temperature is higher than 94 °C in both K<sup>+</sup> and Na<sup>+</sup> conditions for [d(TG<sub>5</sub>T)]<sub>4</sub>.

All these examples highlight the influence of the cation nature and concentration on the stability of G-quadruplexes, where little to no structural changes are expected, and thus the cation dehydration and binding *per se* accounts for most of the stabilization.

**Table 1** Melting temperatures of tetramolecular G-quadruplexes.

sequence	strand concentration ( $\mu$ M)	cation concentration (mM)	$T_m$ K <sup>+</sup> (°C)	$T_m$ Na <sup>+</sup> (°C)	refs
d(TG <sub>3</sub> T)	10	110	48	16	[117]
	200	100	51	—	[273]
	10	110	50	17	[117]
d(T <sub>2</sub> AG <sub>3</sub> )	6400	110	$\approx 60$	—	[274]
d(T <sub>2</sub> AG <sub>3</sub> T)	10	110	55	24	[117]
d(TG <sub>4</sub> T)	10	110	> 94	54.5	[117]
	160	210	> 94	75	[112]
d(T <sub>2</sub> G <sub>4</sub> T <sub>2</sub> )	10	110	> 94	71	[117]
d(AG <sub>4</sub> T)	10	110	> 94	60	[117]
r(UG <sub>4</sub> U)	10	110	> 94	89	[117]
d(TG <sub>5</sub> T)	10	110	> 94	> 94	[117]

## 3.2 General Trends

### 3.2.1 Libraries

The Mergny group contributed several systematic UV-melting based studies on a very large number of oligonucleotides, illustrating the difference of G-quadruplex stabilization by coordination of  $K^+$  or  $Na^+$ . In a 1998 publication illustrating the use of UV-vis spectroscopy to follow the folding of G-quadruplexes, very large differences of melting temperatures in potassium- and sodium-rich conditions were observed for the 26- and 27-mer sequences  $d[T_3A_2G_3(TGTG_3)_3]$  (63 and 37 °C, respectively) and  $d[G_3(TGTGTG_3)_3]$  (62 and 37 °C, respectively) [111]. In a more recent study, the melting temperature of eighty different sequences containing four tracts of three guanines with loops of variable length (between 1 and 15 bases), following the template  $d(G_3L_aG_3L_bG_3L_cG_3)$ , where  $L_a$ ,  $L_b$ , and  $L_c$  are thymines or TTA loops, were determined in presence of 100 mM  $K^+$  or  $Na^+$  [115]. Potassium stabilizes these sequences by on average 18.3 °C more than sodium, but the difference is highly variable since it can be as low as 1.2 °C, and as high as 39.2 °C. The  $T_m$  decreases when the loop length increases both in  $K^+$  and  $Na^+$ , but to different extents – particularly for short loops – as seen from the large  $\Delta T_m$  variability. For loops of 7–15 nt, the difference within a given sequence family tends to be relatively constant, from 1–2 °C (for  $L_a = L_c = TTA$ ) to more than 30 °C (for  $L_a = L_c = T$ ). More generally, there is a strong inverse correlation between total loop length and  $T_m$  for  $K^+$  (each added base leads to a 2 °C drop) but the trend is less clear in  $Na^+$ . Also, the presence of adenines in the loops is favorable in presence of sodium, when the central loop contains at least two nucleotides.

Thirty-six sequences following the general formula  $d(G_3T_3G_3HNHG_3T_3G_3)$  were analyzed in similar conditions (N can be any base, H = C, T or A) [116]. The average difference of  $T_m$  is 12.7 °C in favor of potassium, but is also sequence dependent, ranging from 9.5 °C (for ACC and TGC central loops) to 16.2 °C (for AAT). In the same study, twenty-six additional sequences that vary in length, number of quartets and loop composition and positions were also investigated. The average of (measurable) potassium-sodium  $\Delta T_{ms}$  is 14.4 °C, with a very large sequence dependence, ranging from 7.1 °C ( $d(G_3T_2AG_3CGCG_3T_2AG_3)$ ) to more than 40 °C ( $d(G_3TG_3ACTG_3TG_3)$ ). Overall,  $K^+$  stabilizes particularly well G-quadruplexes containing YDH loops, and poorly ACH loops, while  $Na^+$  favors YDC loops and disfavor ACW loops (Y = C or T; H = A, C or T; W = A or T and D = A, G or T.).

Risitano and Fox have examined a randomized library of oligonucleotides based on the sequence  $d[T(G_3H_2)_3G_3]$ , by FRET-melting, in presence of various concentrations of  $K^+$  and  $Na^+$  (5–100 mM) [94]. A clear increase in  $T_m$  is observed with the concentration, ranging from 27 to 85 °C for  $K^+$ . For  $Na^+$ , lower concentrations (5 and 10 mM) do not allow sufficient folding for the  $T_m$  to be measured, and  $T_{ms}$  at higher concentrations (20–100 mM) range from 24 to 57 °C. The difference of stabilization, where measurable, decreases slightly with increasing cation concentration (29.4–27.6 °C).

Smargiasso *et al.* have investigated the melting temperature of libraries of sequences based on the sequence  $d(G_3W_iG_3W_jG_3W_kG_3)$ , where W is either a thymine or an adenine, and the loop sizes ( $i, j, k$ ) were systematically varied between 1 and 3, yielding a total of 2744 distinct sequences grouped as a function of their loop lengths [39]. The stability in all the groups is higher with 150 mM  $K^+$  than with 150 mM  $Na^+$ , but again this is very sequence-dependent. The stability in  $K^+$  is inversely dependent on loop length: shorter loops promote parallel structures with average  $T_m$  above 80 °C, whereas longer loops (at least one loop of two and one loop of three nucleotides) promote hybrid structures with average  $T_m$  in the 65—70 °C range. Conversely, longer loops yields more stable G-quadruplexes in  $Na^+$ , and the loop length has less influence. Note however that some of these sequences are suspected of forming multimeric G-quadruplexes, notably in  $K^+$  conditions [39].

### 3.2.2 Human telomeric sequences

The human telomeric sequence is certainly the most investigated intramolecular G-quadruplex-forming sequence. Telomeres consist of thousands of tandem repeats of the sequence  $d(T_2AG_3)$ , with a 3'-end overhang of 100—200 nucleotides [275,276]. Telomere sequences are involved in the protection of the ends of chromosome and exhibit similar repeats of guanines in numerous species such as plants ( $d(T_3AG_3)$ ), *Oxytricha* ( $d(T_4G_4)$ ), *Tetrahymena* ( $d(T_2G_4)$ ), or *Bombyx* ( $d(T_2AG_2)$ ).

Włodarczyk *et al.* have measured the melting temperature of the 24-mer  $d[(T_2AG_3)_4]$  sequence in presence of various concentration of alkali metals [85] (Table 2) [85,89,94,110,111,117,118,271]. At ~ 100 mM in  $K^+$ ,  $Na^+$ ,  $Rb^+$ , and  $Li^+$ , the  $T_m$  is respectively 59, 50, 40, and 32 °C, nicely illustrating the important role of alkali metal ions in G4 stability. This ranking is conserved at lower concentrations (~ 50 mM), ranging from 50 to 28 °C, and the G-quadruplex is still reasonably stable at 10 mM in  $K^+$  ( $T_m = 44.3$  °C) and to a lower extent  $Rb^+$  ( $T_m = 26.4$  °C). Higher  $Na^+$  and  $K^+$  concentrations bring further stabilization (62 and 71 °C, respectively, at ~ 300 mM), and partial folding is observed at room temperature with 180—230 mM  $Cs^+$  ( $T_m = 27.4—27.7$  °C). CD spectra were acquired at 2 °C and 20 °C with increasing concentrations of cations. Saturation was not observed for  $Li^+$  at 20 °C and  $Cs^+$  at both temperatures, consistent with their weak G-quadruplex stabilization properties. The presence of high  $Cs^+$  concentrations seemingly destabilizes the G-quadruplex. Conversely, only 4 and 30 mM were enough to reach saturation for  $K^+$  and  $Na^+$ , respectively.

Sugimoto and co-workers have found similar results in 100 mM  $Na^+$  (~ 47 °C) and  $K^+$  (~ 61 °C) [271], and Balagurumoorthy and Brahmachari have observed melting temperatures of 63 and 49 °C in 70 mM KCl and NaCl, respectively [110]. This is somewhat higher than found by Włodarczyk *et al.*, maybe because the values were extracted from 10-data point melting profiles. The 9-mer  $d(G_3T_2AG_3)$ , forming a bimolecular G-quadruplex, was also studied in the same conditions ( $T_m = 42$  and 31 °C, respectively; 20.5  $\mu$ M strand concentration). The 22-mer  $d[AG_3(T_2AG_3)_3]$  was found to be more stable in  $K^+$  than in  $Na^+$  by 7 °C in two distinct publications (62/63 vs. 55/56 °C), and than in  $Li^+$  by 26 °C [89,111].

The difference of stability is identical for the shorter oligonucleotide 21-mer sequence d[(G<sub>3</sub>T<sub>2</sub>A)<sub>3</sub>G<sub>3</sub>], with both the K<sup>+</sup>-form ( $T_m = 65$  °C) and the Na<sup>+</sup>-form being slightly stabilized ( $T_m = 58$  °C) as compared to the 22-mer counterpart [111]. The short 6- and 7-mers d(T<sub>2</sub>AG<sub>3</sub>) and d(T<sub>2</sub>AG<sub>3</sub>T) forming tetramolecular G-quadruplexes are unstable in presence of Na<sup>+</sup> (17 and 24 °C), and much more stable in K<sup>+</sup> (50 and 55 °C) [117].

Modified human telomeric sequences are also under scrutiny. An 18-mer sequence that contains repeats of only two guanines (d[(AG<sub>2</sub>T<sub>2</sub>)<sub>3</sub>AG<sub>2</sub>]) is destabilized as compared to the unmodified 22-mer sequence counterpart, and has almost the same melting temperature in K<sup>+</sup> and Na<sup>+</sup> conditions (42 and 40 °C, respectively) [117]. The addition of extra nucleotides in 5' to obtain the 26-mer d[T<sub>3</sub>A<sub>2</sub>(G<sub>3</sub>T<sub>2</sub>A)<sub>3</sub>G<sub>3</sub>] is also detrimental to the stability but yields a higher  $T_m$  difference (55 vs. 44 °C) [118]. Risitano and Fox have examined the sequence d[T(G<sub>3</sub>T<sub>2</sub>A)<sub>3</sub>G<sub>3</sub>], which contains an additional dT nucleotide in 5' as compared to the non-modified 21-mer sequence, by FRET-melting in presence of various concentrations of K<sup>+</sup> and Na<sup>+</sup> (5—100 mM) [94]. A clear increase in  $T_m$  is observed with the concentration, ranging from 45 to 82 °C in K<sup>+</sup> and 28 to 70 in Na<sup>+</sup>, with the difference of stabilization decreasing from 17 to 12 °C. The high melting temperatures could be attributed to the presence of fluorophores at both ends of the oligonucleotides that may stabilize the G-quadruplexes.

**Table 2** Melting temperature of human telomeric G-quadruplexes.

sequence	cation concentration (mM)	$T_m$ K <sup>+</sup> (°C)	$T_m$ Na <sup>+</sup> (°C)	$T_m$ Rb <sup>+</sup> (°C)	$T_m$ Li <sup>+</sup> (°C)	refs
d[(T <sub>2</sub> AG <sub>3</sub> ) <sub>4</sub> ]	300	71.2	62.2	—	—	[85]
	100	59.0	50.2	40.2	32.0	[85]
	100	61	47	—	—	[271]
	70	63	49	—	—	[110]
	50	50.2	42.4	34.8	27.8	[85]
	10	44.3	—	26.4	23.0	[85]
d[AG <sub>3</sub> (T <sub>2</sub> AG <sub>3</sub> ) <sub>3</sub> ]	110	62	55	—	—	[89]
	110	63	56	—	37	[111]
d[(G <sub>3</sub> T <sub>2</sub> A) <sub>3</sub> G <sub>3</sub> ]	110	65	58	—	—	[111]
d(G <sub>3</sub> T <sub>2</sub> AG <sub>3</sub> ) <sup>a</sup>	70	42	31	—	—	[110]
d(T <sub>2</sub> AG <sub>3</sub> ) <sup>b</sup>	110	50	17	—	—	[117]
d(T <sub>2</sub> AG <sub>3</sub> T) <sup>b</sup>	110	55	24	—	—	[117]
d[(AG <sub>2</sub> T <sub>2</sub> ) <sub>3</sub> AG <sub>2</sub> ]	110	42	40	—	—	[117]
d[T <sub>3</sub> A <sub>2</sub> (G <sub>3</sub> T <sub>2</sub> A) <sub>3</sub> G <sub>3</sub> ]	100	55	44	—	—	[118]
d[T(G <sub>3</sub> T <sub>2</sub> A) <sub>3</sub> G <sub>3</sub> ]	5	45	28	—	—	[94]
	100	82	70	—	—	[94]

<sup>a</sup> bimolecular structure; 20.5 μM strand concentration

<sup>b</sup> tetramolecular structure; 10 μM strand concentration

### 3.2.3 Other sequences

Hardin *et al.* observed in 1991 that the four-repeat *Tetrahymena* telomeric sequence  $d[(T_2G_4)_4]$  is much more stable in  $K^+$  than  $Na^+$  and  $Li^+$  conditions [215]. More recently, Tran *et al.* analyzed the melting temperatures of minimal (*i.e.* without flanking base) telomeric sequences from fourteen species [114]. All the tested sequences bearing three or more consecutive guanines (in particular  $G_3T_{1-4}A$  motifs and  $G_4T_{2,4}$  ciliate motifs) fold into stable G-quadruplexes both in potassium and in sodium. This is also the case of the  $G_2T_2A$  (*e.g.* *Bombyx mori*) and the  $G_2CT_2A$  telomeric motif sequences (*e.g.* *Ascaris lumbricoides*, *Caenorhabditis elegans*). Among the telomeric sequences containing a four repeated  $G_3$  motif, coordination to  $K^+$  induces a stabilization of on average 11 °C as compared to  $Na^+$ , at a 100 mM cation concentration. The difference is the most pronounced for *S. cerevisiae* (68 vs. 50 °C) and the least pronounced for the human telomeres (65 vs. 59 °C). A large difference of stability in  $K^+$  and  $Na^+$  conditions was equally observed in an independent study [111]. *Paramecium* and *L. esculentum* (tomato plant) have degenerated telomeric sequences. The  $T_m$  does not strongly vary among tomato variant sequences, ranging from 55 to 58 °C in NaCl and from 64 to 69 °C in KCl while *Paramecium* variant sequences spanned a broader range: from 52 to 60 °C in NaCl and from 68 to 76 °C in KCl. The ciliates *Tetrahymena* ( $d[(G_4T_2)_3G_4]$ ) and *Oxytricha* ( $d[(G_4T_4)_3G_4]$ ) telomeric sequences containing four repeats of four guanines form particularly stable structure in potassium-rich solutions (> 80 °C), respectively at least 16 and 14 °C more stable than in sodium solutions, which was also observed in previous studies [82,111]. For most of these sequences, and as opposed to tetramolecular G-quadruplexes described earlier, the difference of stability as a function of the cation can be partially ascribed to the formation of distinct topologies (see sections 1.2 and 1.3). Finally, some sequences do not form G-quadruplexes in neither  $K^+$  nor  $Na^+$  (yeasts *S. pombe* and *C. guilliermondii*).

The two-repeat *Oxytricha* telomeric sequence  $d(G_4T_4G_4)$ , which share the same bimolecular, antiparallel topology in  $Na^+$  and  $K^+$  (*vide supra*), has been examined by various techniques. A CD-melting study from 1992 indicates that the 70 mM  $Na^+$ -stabilized G-quadruplex melts cooperatively at 55 °C (20 OD<sub>260</sub> strand concentration) [109]. NMR on a 5—55 °C temperature range revealed that, in presence of  $Na^+$ , unfolding is detected above 35 °C by the heterogeneity of the spectra [107]. Conversely, the spectra acquired with  $K^+$  retain their homogeneity on the full temperature range. These observation were presented as consistent with CD-melting experiments evidencing melting of  $d(T_4G_4)$  in  $Na^+$  at around 40 °C [119], but the comparison is not entirely relevant since (i) such a short sequence certainly forms a G-quadruplex of distinct topology (parallel), as seen from the CD signature, (ii) of distinct molecularity (tetramolecular) and strand concentration (2.2 vs 0.1 mM, respectively), and (iii) the cation concentration is different (50 vs 200 mM NaCl) (see section 3.1). Incidentally, this latter study by Kallenbach *et al.* shows that the addition of a single 3'-dT nucleotide dramatically increases the melting temperature (+ 20 °C), and that that the G-quadruplex formed in 200 mM KCl is not entirely melted at 100 °C [119].

Petraccone *et al.* have used DSC to determine the  $T_m$  of the two-repeat *Oxytricha*

telomeric sequence forming a bimolecular G-quadruplex  $[d(G_4T_4G_4)]_2$ , in 200-mM sodium solutions [112]. Analysis of the heating curve revealed a two-step dissociation, centered around 47 and 67 °C, contrary to what has been proposed in earlier studies [109,119]. The low-temperature transition has been attributed to an intramolecular pre-melting event. FRET-melting experiments were performed on the four-repeat  $d[T(G_4T_4)_3G_4T]$  sequence, but the melting temperatures were very dependent on the heating/cooling rates, and the melting curves presents an hysteresis, because of the exceptionally slow dissociation rates observed in  $K^+$  and  $Na^+$  conditions (see 3.3) [249]. Overall, higher  $T_m$  were observed with  $K^+$  than  $Na^+$  and with increasing cation concentrations. In some cases, biphasic melting profiles were also observed.

In various studies, TBA ( $d(G_2T_2G_2TGTG_2T_2G_2)$ ) was shown to be unfolded in  $Li^+$ , fairly unstable in 100 mM NaCl ( $T_m \approx 20$  °C), and largely more stable in  $K^+$  (48—50 °C) [89,98,111]. Large differences were also observed for c-myc sequences in FRET-melting experiments, notably for  $d(G_4AG_4TG_4AG_4)$  ( $\Delta T_m = 25$  °C at 5 mM in  $K^+$  vs.  $Na^+$ ) [94]. The melting temperatures of the closely related sequences  $d[G_3(T_2G_3)_3]$  and  $d[TG_3(T_2G_3)_3T]$  in presence of various alkali cations, ranging from 85 and 77 °C, respectively, in 100 mM  $K^+$  to less than 18 °C in 100 mM  $Li^+$  [208]. An intermediate behavior is found for  $Na^+$  and  $Rb^+$  that give comparable melting temperatures for both sequences, around 45 °C.

### 3.2.4 Summary

Overall, cation stabilization follows the order  $K^+ > Na^+, Rb^+ > Cs^+ > Li^+$ . From all these studies it can clearly be seen that the difference of G-quadruplex stability between  $K^+$  and  $Na^+$  conditions is always in favor of potassium. However, one cannot predicts the stability of G-quadruplex in sodium from the result obtained in potassium or *vice versa*, since this difference spans from a couple of degrees to almost 40 °C, depending on the sequence. Differences in stability are particularly important for tetramolecular assemblies ( $[TG_3T]_4$ ,  $[TG_4T]_4$ , single repeat human and *Oxytricha* telomeric sequences), where the structure is the same regardless of the cation. Increasing cation concentrations lead to higher  $T_m$ , as expected from the law of mass action, and reduces the  $T_m$  difference between cations. This difference is classically attributed to the greater cost of  $Na^+$  dehydration, which is only partially compensated by a more favorable free energy of binding than  $K^+$  [88,105]. However, the differences in  $T_m$  are also strongly sequence – and hence sometimes structure – dependent. The influence of cations on the structure of G-quadruplexes is addressed in section 4.

## 3.3 Kinetics of strand association, dissociation, folding and unfolding

Early work on tetramolecular G-quadruplex nucleic acids show that they generally display slow association and dissociation rates [277,278]. As mentioned previously, the guanine imino proton exchange is very slow as compared to duplex DNA, and lies in the day-to-week

scale in D<sub>2</sub>O solutions for the tetramolecular RNA G-quadruplex [r(UG<sub>4</sub>U)]<sub>4</sub> [279]. Intermolecular G-quadruplex formation is extremely slow, and the rate is very concentration dependent (for both the strand and the cation). A key study regarding the association and dissociation of short G-rich strands in tetramolecular G-quadruplex, in Na<sup>+</sup> or K<sup>+</sup> conditions, demonstrated that this association is close to fourth order in monomer, whereas the dissociation is first order in G-quadruplex [117]. The association rate constant decreases with increasing temperatures, reflecting a negative apparent activation energy, and is increased by an increase in strand or alkali cation concentration. On the other hand, the dissociation is also temperature dependent, but is virtually not affected by the ionic strength. The association half-time of [d(TG<sub>4</sub>T)]<sub>4</sub> ranges from six seconds to more than a hundred years for strand concentrations decreasing from 1 mM to 1 μM, in 110 mM Na<sup>+</sup>. Substituting Na<sup>+</sup> by K<sup>+</sup> affords a 20—50 fold increase in association constants, fairly constant in a 2—37 °C temperature range and 50—300 mM cation concentration range. Although an increase in sodium concentration from 50 to 400 mM does not affect significantly the melting temperature of [d(TG<sub>4</sub>T)]<sub>4</sub> (see section 3.1), it dramatically increases the association rates. An increase of one order of magnitude of Na<sup>+</sup> concentration leads to 1000—2000 increase in association constant. The fact that the cation-dependent increase in stability observed for many G-quadruplexes is mainly reflected in the association rate constant is similar to duplex and triplex nucleic acids (chapter 6 by Ennifar *et al.*?). The dissociation is not observed in most cases for G<sub>5</sub> and longer tracts in Na<sup>+</sup>, and G<sub>4</sub> and longer tracts in K<sup>+</sup>. Melting and temperature-jump experiments of fluorescently-tagged G-quadruplexes ([d(TG<sub>4</sub>T)]<sub>4</sub>, [d(G<sub>4</sub>T)]<sub>4</sub>, [d(TG<sub>4</sub>)]<sub>4</sub>, [d(G<sub>4</sub>)]<sub>4</sub>) at different heating/cooling rates in Na<sup>+</sup> solutions confirmed that the rates of annealing are very slow [236].

Another comprehensive study on the cation-dependent formation of DNA [d(TG<sub>*n*</sub>T)]<sub>4</sub> (*n* = 3—6) and RNA [d(UG<sub>4</sub>U)]<sub>4</sub> tetramolecular G-quadruplexes has been provided by Bardin and Leroy [164]. NMR experiments demonstrated that increasing salt concentrations favors the association rate, but that the cation nature (K<sup>+</sup>, Na<sup>+</sup>, Li<sup>+</sup>) only weakly contributes. However, the lifetime of the G-quadruplex is much higher in K<sup>+</sup> than in Na<sup>+</sup> and Li<sup>+</sup>, and increases with the guanine tract length. The authors suggest that, in 125 mM K<sup>+</sup> solutions, the G-quadruplex formation proceeds step-by-step via successive duplex and – kinetically limiting – triplex intermediates, and follows a reaction order of three. On the other hand, the reaction order is found to be four in Li<sup>+</sup>, and each step is strand-concentration dependent. The formation of mismatched G-quadruplexes, particularly at low temperature and K<sup>+</sup> concentrations, is believed to impede the formation of the canonical G-quadruplex by kinetic trapping. In the same vein, Kim *et al.* have performed PAGE experiments showing that the formation of a tetramolecular RNA G-quadruplex by the r(GC<sub>2</sub>GAUG<sub>2</sub>UAGUGUG<sub>4</sub>U) sequence is faster in Na<sup>+</sup> than K<sup>+</sup>, despite the higher stability observed in the latter case [272]. They also suggest that the formation of unproductive intermediates impedes the formation of the G-quadruplex in potassium conditions.

The cation-dependent kinetics of intramolecular G-quadruplex formation by telomeric sequences was investigated by different groups. Raghuraman and Cech evidenced that the half-life of the *Oxytricha* telomeric sequence [d(G<sub>4</sub>T<sub>4</sub>)]<sub>4</sub> unfolding is shorter in 50 mM Na<sup>+</sup>

than  $K^+$  (4 vs 18 hours at 37 °C) [280]. The FRET-tagged sequence d[T(G<sub>4</sub>T<sub>4</sub>)<sub>3</sub>G<sub>4</sub>T] shows exceptionally slow dissociation rates in  $K^+$  and  $Na^+$  conditions [249]. The half-life at 37 °C is roughly 10 years in 50 mM  $K^+$ , while the corresponding association half-life is of about 0.9 seconds. The rate of association is 10—60 fold faster than in sodium in the 50—100 mM cation concentration range. Differences in dissociation rates are much more marked ( $t_{1/2}^{37\text{ °C}} = 3900\text{ s}$  in 50 mM  $Na^+$ ). From these results, Fox and coworkers suggested that, although there are differences in the rates of association of G-quadruplexes under different conditions, the dissociation rate constant is the most important factor that affects the relative stabilities [249], in clear contrast with the abovementioned conclusions of Mergny *et al.* on tetramolecular G-quadruplexes [117]. However, as mentioned in the latter study, and others [164,209], association rates decrease with increasing temperature and suggests a nucleation-zipper mechanism for G-quadruplex formation. Using a PNA trap, large cation-dependent differences were observed for the average unfolding time constants of the human telomeric sequence d[G<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>], in presence of either 100 mM  $K^+$  (40 h at 20 °C) or  $Na^+$  (30 min) [281]. This is qualitatively consistent with the higher melting stability observed in  $K^+$  (see section 3.2.2).

Time-dependent FRET monitoring as a function of complementary strand hybridization of TBA revealed a first-order kinetics, where the unfolding of the G-quadruplex is the rate-determining step [98], consistent with some previously mentioned studies [280,281]. The unfolding rate is also strongly cation-dependent ( $62 \times 10^5$  and  $3200 \times 10^5\text{ s}^{-1}$  at 10 °C, in respectively  $K^+$  or  $Na^+$ ). If the nature of the cation affects largely the dynamics of G-quadruplex folding and unfolding, other factors are to be taken into account such as the cation concentration, and the oligonucleotide sequence. At low  $K^+$  concentration, the human telomeric sequence display a relatively fast dynamic behavior (< 100s) [243]. The sequence – and hence structure – that is studied also affects the dynamics of folding and unfolding. As seen above, the *Oxytricha* telomeric sequence can maintain its G-quadruplex structures for longer times than its human counterpart [249,280], even at low  $K^+$  concentrations [243], likely because it contains four quartets instead of three.

In the past years, several groups have explored folding intermediates and their interactions with cations (mostly  $K^+$ ). G-quadruplex folding intermediate may adopt pre-organized structures inclined to fold into a G-quadruplex after cation binding. Mashimo and coworkers have investigated the folding pathways of the human telomeric sequence [282-284]. A combination of *ab initio* calculations and molecular dynamics hinted at the formation of hairpin containing Hoogsteen G•G base pairs, possibly yielding a more stable triplex intermediate. Coordination of  $K^+$  favors each step of the folding by decreasing the electrostatic repulsions [284]. Binding to a hairpin near the second lateral TTA loop was found to be preferable, and more generally to 5'-G(*syn*)G-(*anti*)-3'/5'-G(*syn*)G(*anti*)-3', rather than to 5'-G(*anti*)G(*anti*)-3'/5'-G(*syn*)G(*syn*)-3'.

The Chaires group also thoroughly investigated the folding and unfolding of human telomeric sequences in presence of  $K^+$ , notably by CD, 2-aminopurine fluorescence, FRET, and molecular modeling (molecular dynamics) [285-287]. The latest results suggest a four-step pathway [285] (Figure 6A) [171,285,288]. The first folding intermediate I1 is a mixture

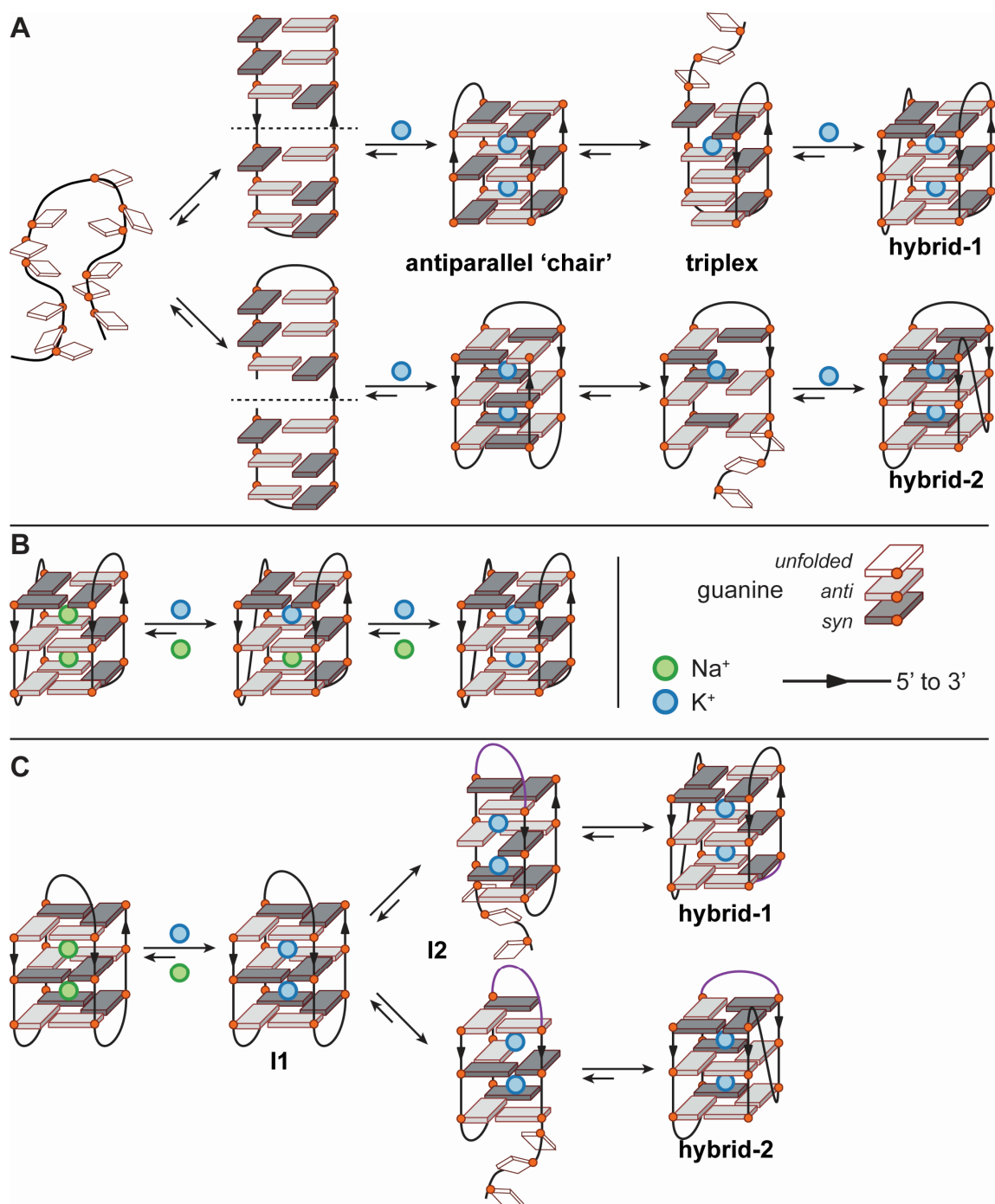


of hairpin structures that is rapidly formed thanks to cation-induced collapse of the unfolded strands. Intramolecular fold over of I1 may form antiparallel G-quadruplex structures I2 upon binding to  $K^+$  ( $\tau_1 \approx 0.1$  s in 25 mM KCl, at 25 °C), the latter further converting to a triplex intermediate I3 ( $\tau_2 \approx 3700$  s) that finally yields the final G-quadruplex ( $\tau_3 \approx 750$  s). Other groups have suggested the triplex structure as an unfolding intermediate for the human telomeric sequence [289-291], including in conformational switching context [165]. Zhang and Balasubramanian stated that, regardless of the structure of the DNA or RNA quadruplex formed, the folding pathway involves two steps initiated by the binding of a single  $K^+$  cation, leading to a hairpin intermediate, followed by the formation of a triplex [292].

A triplex folding intermediate, stabilized by a  $K^+$  cation, was also proposed for the folding of TBA [293,294]. In MD calculations,  $K^+$  was manually positioned between the two triads according to distances observed for the high-resolution crystal structure of TBA [161]. The guanine G10, despite being out of plane, seems well positioned to coordinate  $K^+$  with its O6.  $Na^+$  poorly stabilizes the triplex of TBA (folding is partially visible by NMR under 1 °C), which is not surprising as it was shown to be a poor TBA stabilizer [89,98,111].

Plavec *et al.* have recently proposed a new folding intermediate for the G-quadruplex of the *Oxytricha* sequence Oxy-1.5, namely i-Oxy-1.5, existing as two symmetric bimolecular forms containing G•G N1-carbonyl symmetric base pairs [246]. This intermediate was characterized by NMR, TDS, CD, DSC, and PAGE, nicely illustrating how the use of a combination of analytical methods allows to unearth insightful results. DOSY, PAGE and concentration-dependent  $T_m$  results demonstrated that i-Oxy-1.5 is bimolecular and has a comparable compactness than the final G-quadruplex. It however does not contain tetrads as seen by NMR and DSC, the latter allowing to evaluate the difference in enthalpy of unfolding between i-Oxy-1.5 and Oxy-1.5 at around 48 kcal.mol<sup>-1</sup>. The three-fold lower enthalpy observed for the intermediate has been attributed to twice as less H-bonds, less favorable stacking interactions, and, more importantly, the absence of cation coordination. Its formation is a slow kinetically governed process, while  $K^+$  binding and subsequent restructuring consist of two fast processes.

Studies investigating folding intermediates mostly focus on  $K^+$ , and hence do not give much information about the cation nature dependence in the folding process. Comparable behaviors may be expected as compared to final G-quadruplex structures (*e.g.*  $Na^+$  would stabilize the intermediates to a lower extent), as hinted by the triplex intermediate of TBA [294], but insufficient data had been gathered so far. However, they all highlight the rapidity and the importance of processes that involve cation binding as compared to strand folding and unfolding. This will be further emphasized in sections 3.4 and 5 that focus on cation exchange and conformational switching.



**Figure 6** A) Folding of the human telomeric sequence in K<sup>+</sup> conditions as proposed in reference [285]. B) Step-wise Na<sup>+</sup>/K<sup>+</sup> exchange mechanism of the human telomeric sequence proposed in reference [171]. C) Na<sup>+</sup>/K<sup>+</sup> exchange and structural interconversion proposed in reference [288].

### 3.4 Cation Exchange Mechanisms

$\text{Na}^+$  and  $\text{K}^+$  have a different mobility inside G-quadruplexes. Sodium is moving faster [295], presumably because its size does not hamper passing through tetrads [97]. The mobility of  $\text{Na}^+$  depends on the binding site, as observed by  $^{23}\text{Na}$  NMR spectroscopy [190,257], and  $^{15}\text{NH}_4^+$  displacement NMR experiments [97]. In the bimolecular G-quadruplex formed by  $\text{d}(\text{G}_4\text{T}_4\text{G}_4)$  (see section 1.3 and Figure 5),  $\text{Na}^+$  ions bound in loops have a residence lifetime of 220  $\mu\text{s}$  at 15 °C ( $k_{\text{off}} = 4.5 \times 10^3 \text{ s}^{-1}$ ) [190,257], whereas the lifetime in the central stem is at least two-order of magnitude higher [97]. Sodium is in a relatively dynamic exchange between coordination sites and the bulk medium, and the more affine potassium can displace sodium from these binding sites [256,257]. Replacement of  $\text{Na}^+$  by  $\text{K}^+$  within the G-quadruplex central stem is completed within  $\sim 250 \mu\text{s}$  [190,257,296].

Hud *et al.* suggested that two  $\text{K}^+$  binding events occur on the G-quadruplex  $[\text{d}(\text{G}_3\text{T}_4\text{G}_3)]_2$ , where the displacement of  $\text{Na}^+$  by increasing amount of  $\text{K}^+$  does not impact significantly the topology. Hence, each sodium cation is exchanged in a step-wise process where the intermediate is a mixed-cation species that is in fast exchange (on the NMR time-scale) with the pure species [88]. A decade later, Plavec and co-workers have shown that the conversion of the  $^{15}\text{NH}_4^+$ -coordinating  $[\text{d}(\text{G}_3\text{T}_4\text{G}_4)]_2$  G-quadruplex to the  $\text{K}^+$ -form also proceeds by step-wise exchange of the cations [260]. The second binding site is not exchanged before the first one is fully occupied by potassium, yielding a mixed-cation intermediate. This can be explained by the differences between the two coordination sites; the variations in  $\Delta G$  between the two cations are not identical for both sites ( $\Delta\Delta G = -5.7$  and  $-4.3 \text{ kcal.mol}^{-1}$ , respectively, including cation dehydration energies).  $^{23}\text{Na}$  NMR spectroscopy also evidenced that increasing amount of  $\text{K}^+$  displaces  $\text{Na}^+$  cations from the abovementioned  $[\text{d}(\text{G}_4\text{T}_4\text{G}_4)]_2$  G-quadruplex, without unfolding of the G-quadruplex, and that both cations may reside in the G-quadruplex simultaneously [190]. Differences in binding sites leading to mixed-cation species were also observed with quadruplexes built from lipophilic guanosine derivatives [297]. Within a four tetrad construct, the replacement of the central  $\text{Na}^+$  by  $\text{K}^+$  takes place before the outer cations are exchanged, whereas the wider  $\text{Cs}^+$  cation can exclusively be coordinated by the central binding site.

The  $\text{Na}^+$  to  $\text{K}^+$  exchange in the human telomeric was also particularly intensely investigated. It was proposed on the basis of NMR and CD experiments that human telomeric sequences (26 and 22-mers) form an antiparallel ‘basket-type’ structure in  $\text{Na}^+$  solutions and a hybrid-1 structure in  $\text{K}^+$  [165], so here the cation exchange is accompanied by a dramatic change in G-quadruplex topology. Therefore the authors hypothesized that the cation exchange from the former to the latter involves a partial unfolding and restructuration steps, after the cation exchange (see section 5). In the same vein, Gray *et al.* have shown that the (30 mM)  $\text{Na}^+$  to (50 mM)  $\text{K}^+$  exchange of the 22-mer  $\text{d}[\text{AG}_3(\text{G}_3\text{T}_2\text{A})_3]$  is also followed by a conformation change, from the  $\text{Na}^+$ -basket type to the  $\text{K}^+$ -hybrid (Figure 6C) [288]. Three distinct kinetic processes and two intermediates I1 and I2 are involved (see section 5 for more details). The cation exchange per se is believed to be fast (5 ms), as compared to the ensuing conformational change (11—14 min). Chang *et al.* have proposed that the 23-mer sequence

d[TAG<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>] adopts the same hybrid-1 topology in both K<sup>+</sup> and Na<sup>+</sup> solutions despite very different CD spectra, and thus exclude unfolding of the G-quadruplex for the exchange to take place (Figure 6B) [171]. The exchange is complete in around 80 seconds, and its mechanism involves two steps, each Na<sup>+</sup> cation being exchanged independently *via* a mixed-cation intermediate as observed by aforementioned studies on bimolecular G-quadruplexes [88,190,260]. This step-wise exchange is proposed notably on the basis of a 22-fold difference between the binding constants of the two coordination sites.

The residence lifetimes of Na<sup>+</sup> cations are much shorter than tetrad breathing movements. This strongly suggests that Na<sup>+</sup> ions exchange through the central stem of G-quadruplexes without requiring partial tetrad opening [298]. Accordingly, all studies show that the exchange of Na<sup>+</sup> by K<sup>+</sup> does not require unfolding, although unfolding may take place after cation exchange to allow conformational switching (see section 5). Consistent with that hypothesis is the lower residence lifetime measured for tetramolecular G-quadruplexes vs. mono- and bi-molecular G-quadruplexes, whose loops could impede cation release [189].

Solution-state NMR experiments on <sup>15</sup>NH<sub>4</sub><sup>+</sup> evidenced that cation movement is slower at the 5'-end of G-quadruplexes, and that it is slower through an all-*syn* tetrad than through all-*anti* one [299]. NMR also confirmed that cation movement occurs between each coordination site and the bulk medium rather than between the coordination sites. Exchange of “cation 1” by “cation 2” can thus be described as four successive steps: (i) release of cation 1, (ii) hydration of cation 1, (iii) dehydration of cation 2, and (iv) coordination of cation 2 [62]. The highest energetic cost resides in step (i) where the coordination of the cation shall be removed, and, as a direct consequence, the structure of the G-quadruplex might change. Molecular dynamics simulations suggest that the entering or releasing of K<sup>+</sup> from the G-quadruplex is accompanied by significant changes in O6 distances [295]. Li<sup>+</sup> induces a shrinking of the structure, whereas larger cations (Rb<sup>+</sup>, Cs<sup>+</sup>) significantly deform the structure and hence destabilizes the G-quadruplex. Re-hydration of cation 1 (step ii) is, however, favorable. The following steps (iii) and (iv) are the reverse mechanism [62]. Note however that the final G-quadruplex can exhibit a different structure than the starting one, which would affect the net energetic change. Therefore, this net energy change (2±0.5 kcal.mol<sup>-1</sup> for a three-tetrad system [91]; see section 1.3) is a result of solvation, coordination, and structuration terms.

Reshetnikov *et al.* have described the cation coordination process with TBA, thanks to molecular dynamics simulation, and hybrid quantum mechanics/molecular mechanics simulations [300]. Cations enter TBA through ‘gates’, formed by either the G8 base at the base from the central TGT loop and the space between guanines’ O6 from the upper tetrad, or by the non-canonical T–T pair and the space generated by guanine’s O6 from the lower quartet. When the cation is captured by the quadruplex, it enters rapidly into the central stem, and the gates tend to close (the initial binding event lies in the ns range for both Na<sup>+</sup> and K<sup>+</sup>). Conversely, in absence of coordinated cations, the gates are open. The TGT loop modulates the cation binding process in three ways, by (i) slowing down the cation binding by obstructing the gates, (ii) contributing to keep the coordinated cation in the quadruplex, and (iii) avoiding the quadruplex to collapse when no cation is bound. During the cation exchange

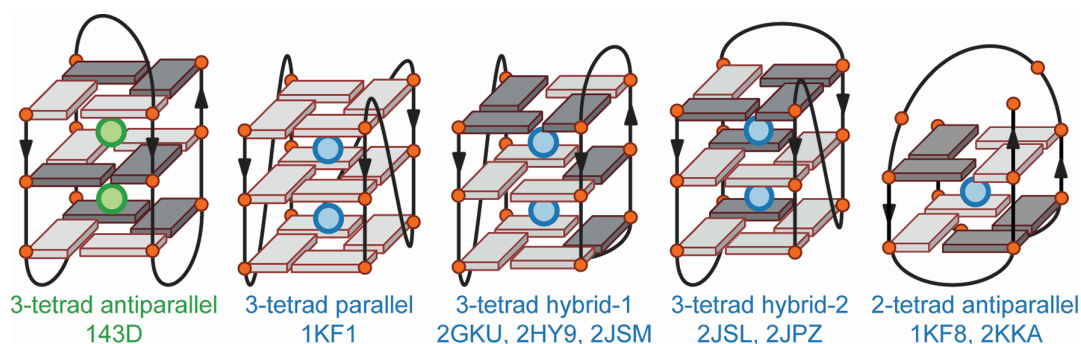
between the bulk medium and other cations already coordinated, the release of the bound cation is correlated with the initial binding of the incoming cation. In short quadruplexes (small number of tetrads), the initial binding of a new cation can lead to electrostatic repulsions with the central stem-coordinated cation, eventually leading to the release of the latter.

Recently, DFT-D3 calculations have evidenced that binding of a second ion in a three-tetrad system is accompanied by a large overestimation of the cation-cation repulsion at the molecular mechanics level of experiment because of its lack of polarization term [202], leading to large energy differences with the quantum mechanics level (20 kcal.mol<sup>-1</sup> in the gas phase). Indeed, the polarization effect between the central stem and the cations decrease the energy cost of coordination of the second cation, by reducing the electrostatic repulsions of the cations. Because of this bias in the force fields that are currently employed, the authors question the validity of MD simulations published so far regarding the treatment of cations (reduced coordination rate, frequent cation ejections, destabilization of the cation coordination in loops above tetrads), and future work in this direction is therefore needed.

## 4 Influence of alkali metal ions on G-quadruplex structures

### 4.1 Case Study: The Human Telomeric G-Quadruplex Sequence

We have seen in section 3.2.2 that the G4 community has a predominant interest in the human telomeric sequence. This particular attention is part historical, with a number of seminal studies appearing in the late eighties and early nineties [81,82,301], part therapeutic [302-304], but also arises from its particularly intriguing polymorphism [172]. As described below, this polymorphism depends on a number of factors, including the cation nature, the precise sequence (number of repeats, flanking sequences), the putative presence of co-solvent, and other experimental factors imposed by the chosen analytical method. Many topologies have been proposed based on various methodologies including NMR [25,27-32,50,168,170-174,242], X-ray [26,150], ESI-MS [231,305], CD [85,173,174,245,306-308], single molecule FRET [233], native [271] or denaturing PAGE [237-239], <sup>125</sup>I-radioprobng [239,240], UV-melting [271], fluorescence spectroscopy [242], Raman scattering [245], analytical ultracentrifugation and molecular modeling [203]. Available high-resolution structures of 4-repeat human telomeric G-quadruplexes are depicted in Figure 2 and schematized in Figure 7 [25-32].



**Figure 7** Topologies of the four-repeat DNA human telomeric G-quadruplex deposited in the PDB [25-32].

#### 4.1.1 The Intramolecular Folding of dAGGG(TTAGGG)<sub>3</sub>

The 22-mer sequence d[AG<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>] containing four guanine repeats and a single flanking nucleotide (5'-dA) has been particularly studied. A first structure has been solved by NMR, in Na<sup>+</sup>-rich conditions, as early as 1993 by Wang and Patel (PDB ID: 143D) [25]. The structure formed is an antiparallel “basket-type” G-quadruplex, characterized by three *syn•syn•anti•anti* tetrads, with each G-tract following the pattern 5'-*syn-anti-syn*, connected by consecutive lateral-diagonal-lateral loops.

Contrastingly, in K<sup>+</sup> conditions, several different structures were found, highlighting the influence of the cation on G-quadruplex structures, but also, regrettably, of the structural analysis method. Parkinson *et al.* published the crystal structure of d[AG<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>] in 2002 (PDB ID: 1KF1) [26]. It exhibits an all parallel topology, with three tetrads formed by *anti* guanines. The K<sup>+</sup> cations are positioned between the tetrads at equidistance from each guanine O6 (2.7 Å), and separated from each other by ~ 3.4 Å (G-quartet rise: 3.1 Å). If this structure differs significantly from the solution-based Na<sup>+</sup> structure described above, it also differs from more recent solution-based structure obtained with K<sup>+</sup> samples.

d[AG<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>] in K<sup>+</sup> solutions actually adopts a mixture of antiparallel and hybrid structures [309,310]. <sup>125</sup>I radioprobng suggested that an antiparallel fold is present in both Na<sup>+</sup> and K<sup>+</sup> conditions, although in potassium the topology would be chair-type (three lateral loops) rather than basket-type, and in a mixture with other topologies [239]. Vorlícková *et al.* have argued that the solution structure has essentially the same antiparallel topology with both cations, based on CD experiments [306], contrary to earlier reports by three other groups that support an antiparallel-to-hybrid switch upon addition of K<sup>+</sup> to a Na<sup>+</sup>-solution [95,165,311]. The conversion from the Na<sup>+</sup> form to the K<sup>+</sup> form could be too fast to reflect a structural change as important as reported previously, and the differences in CD signatures could be assigned to changes in tetrad stacking due to specific K<sup>+</sup> coordination.

A similar observation (change in CD signature but no change in topology) was proposed by Chang *et al.* for the 23-mer d[TAG<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>] [171]. In the same vein, Bombard *et al.* have observed an identical platination site (namely dA13) in both Na<sup>+</sup> and K<sup>+</sup> conditions, by gel analysis of 3'-exonuclease-digested d[AG<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>], previously incubated with platinum

complexes [238]. An earlier study involving *cis*- and *trans*-platinum complexes yielded similar results [237]. It is thus possible that the basket type is at least partially populated in  $K^+$  conditions, and binding to this structure would be favored by the platinum complexes. Alternatively, different structures such as the 2-tetrad antiparallel fold could lead to identical platination sites.

It is likely that the crystal structure of Parkinson *et al.* is influenced by crystal packing forces [26]. Incidentally, water-depleting co-solvents such as polyethylene glycols [307,312], ethanol [306], and acetonitrile [309] promote the parallel fold. However, other groups have stated that this fold is not the favored one in physiological conditions [312-314]. Hänsel *et al.* have suggested by a combination of NMR and fluorescence spectroscopy that d[AG<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>] predominantly adopts the hybrid-1 conformation *in vivo*, *ex vivo*, and in dilute potassium-based solution, and confirm the observation of a parallel fold in water-depleted conditions [242]. Renciuk *et al.* also suggested that the difference between X-ray and NMR experiments arises from the unexpected DNA concentration dependence on the human telomeric intramolecular G-quadruplex [306]. Abu-Ghazalah and coworkers have observed a mixture of antiparallel and parallel structures at high strand concentrations (2 mM), in  $Na^+$  solutions [245]. The parallel topology was tentatively attributed to either higher-order structures formed by propeller monomer, or tetramolecular assemblies.

#### 4.1.2 Other Human Telomeric Sequences

The crystal structure of the 12-mer d[(TAG<sub>3</sub>T)<sub>2</sub>] is a parallel in  $K^+$ , bimolecular G-quadruplex [26], whereas the solution structure is a mixture of parallel and antiparallel bimolecular G-quadruplexes in solution [170]. Bolton *et al.* suggested that the  $Na^+$  fold is antiparallel [308]. The RNA human telomeric (TERRA) counterpart r[(UAGGGU)<sub>2</sub>] does not exhibit the same cation-dependency since it associates into a bimolecular ‘propeller’ G-quadruplex in  $K^+$ -crystals [150], and in both  $Na^+$  [173] and  $K^+$  solutions [174], reminiscent of the DNA crystal structure albeit for changes in sugar puckering. The topology of RNA 4-repeat sequences do not exhibit any cation-dependency either. CD suggests that r[(UUAGGG)<sub>4</sub>] forms a parallel G-quadruplex in  $Na^+$  [173] and  $K^+$  solutions [174]. In the absence of 5'-flanking nucleotides, the bimolecular G-quadruplex [r(G<sub>3</sub>U<sub>2</sub>AG<sub>3</sub>U)<sub>2</sub>] further dimerizes by 5'-5' stacking in  $K^+$  solutions, as seen by NMR [50]. This was also observed on the 12-mer by ESI-MS, but with ammonium ions instead of alkali cations [224].

Unlike DNA 12-mers, no structure difference was observed for the 16-mer three-repeat d[G<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>2</sub>T] in  $Na^+$  and  $K^+$ , when associating into an asymmetric, antiparallel bimolecular G-quadruplex [168,172]. The bimolecular G-quadruplex was also observed in  $Na^+$  by association of two three-repeat strands, where one provides only one G-tract [168].

As stated earlier, four-repeat human telomeric sequences fold into a variety of structures. Vorlícková *et al.* have proposed that the 21-mer d[G<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>] folds into an antiparallel G4 in both  $Na^+$  and  $K^+$ , alike d[AG<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>] [306]. Other groups have attributed the CD signature to a mixture of hybrid and antiparallel folds [307]. Similar conclusions were

drawn from  $^{125}\text{I}$ -radioprobings experiments, that also suggest that the presence of 5'-flanking nucleotides stabilizes an hybrid fold in  $\text{K}^+$ , while  $\text{Na}^+$  promotes an alternative basket structure [240]. NMR experiments show that the addition of 5' and 3' flanking sequences stabilizes hybrid [27-30,171] (PDB ID: 2GKU, 2HY9, 2JSM, 2JSL, 2JPZ), or 2-tetrad antiparallel topologies (2KF8, 2KKA) [31,32], in  $\text{K}^+$  solutions, which are all distinct from the  $\text{Na}^+$  basket type of  $\text{d}[\text{AG}_3(\text{T}_2\text{AG}_3)_3]$  (143D) (Table 3) [25-32,50,150,168,170-174,271,306-308]. Hybrid-1 and Hybrid-2 structures are characterized by successive double chain reversal-lateral-lateral and lateral-lateral-double chain reversal loops, respectively. Chang *et al.* have used NMR to show that  $\text{d}[\text{TAG}_3(\text{T}_2\text{AG}_3)_3]$  adopts the same hybrid-1 topology in both  $\text{K}^+$  and  $\text{Na}^+$  solutions, despite different CD signatures [171]. A number of sequences starting by a guanine, including the 21-mer  $\text{d}[\text{G}_3(\text{T}_2\text{AG}_3)_3]$ , adopt the 2-tetrad antiparallel topology (sometimes coined 'form 3'), that contains 2 tetrads, with lateral-diagonal-lateral loop bases being involved in externally stacked triplets/base pairs [31]. Gabelica *et al.* have evidenced by native ESI-MS experiments the binding of a single  $\text{K}^+$  cation at low KCl concentrations, consistent with the 2-tetrad folding, while higher concentrations lead to the binding of a second  $\text{K}^+$ , either by binding between an external tetrads and loop bases, or by conversion to a 3-tetrad hybrid structure [229]. Włodarczyk *et al.* have acquired CD spectra of  $\text{d}(\text{T}_2\text{AG}_3)_4$  at 2 and 20 °C at increasing concentrations of alkali cations [85]. No conclusion was drawn regarding the relative topologies formed, but the  $\lambda_{\text{max}}$  were reported (290 nm for  $\text{K}^+$ , 295 nm for  $\text{Na}^+$  and  $\text{Rb}^+$ , 300 nm for  $\text{Cs}^+$ , and 301 nm for  $\text{Li}^+$ ).

Renciuk *et al.* have suggested that, although flanking nucleotides can stabilize hybrid topologies, longer telomeric sequences fold in an antiparallel fashion [306]. The use of site-specifically  $^{15}\text{N}$  labeled G4-units in native-like single stranded telomeric in high resolution NMR experiments revealed that the 3'-terminal and internal G4 unit predominantly coexist in 2-tetrad antiparallel basket and hybrid-2 structures, arranged in "beads-on-a-string"-like fashion [242]. Sugimoto and co-workers have suggested that long sequences (5—12 repeats) form intramolecular G-quadruplexes arranged in distinct G-quadruplex units connected by TTA linkers in a beads on a string fashion, in both  $\text{K}^+$  and  $\text{Na}^+$  solutions, based on UV-melting experiments [271].  $\text{Na}^+$  promotes the antiparallel basket topology, whereas potassium is believed to lead to a mixture of parallel and antiparallel structures, or hybrid structures.



**Table 3** Structures adopted by selected DNA and RNA human telomeric sequences

Sequence (5' to 3')	Cation	Topology	Assay	PDB ID	Reference
d[(TAG <sub>3</sub> T) <sub>2</sub> ]	Na <sup>+</sup> (140 mM)	dimeric antiparallel	CD	_____	[308]
	K <sup>+</sup> (100 mM) <sup>a</sup>	dimeric parallel + antiparallel	NMR	_____	[170]
	K <sup>+</sup>	dimeric parallel	X-ray	1K8P	[26]
r[(UAGGGU) <sub>2</sub> ]	Na <sup>+</sup> (215 mM) <sup>b</sup>	dimeric parallel	NMR	_____	[173]
	K <sup>+</sup> (100 mM) <sup>a</sup>	dimeric parallel	NMR	2KBP	[174]
	K <sup>+</sup>	dimeric parallel	X-Ray	3IBK	[150]
r(G <sub>3</sub> U <sub>2</sub> AG <sub>3</sub> U)	K <sup>+</sup> (100 mM) <sup>a</sup>	5'-5' stacked parallel dimer	NMR	2M18	[50]
d[G <sub>3</sub> (T <sub>2</sub> AG <sub>3</sub> ) <sub>2</sub> T]	Na <sup>+</sup> (140 mM) <sup>c</sup>	dimeric antiparallel	NMR	2AQY	[168]
	K <sup>+</sup> <sup>d</sup>	dimeric antiparallel	NMR	_____	[172]
d[G <sub>3</sub> (T <sub>2</sub> AG <sub>3</sub> ) <sub>3</sub> ]	Na <sup>+</sup> (145 mM)	antiparallel (basket)	CD	_____	[306]
	K <sup>+</sup> (variable) <sup>e</sup>	antiparallel (basket)	CD	_____	[306]
	K <sup>+</sup> (50 mM)	hybrid mixture	CD	_____	[307]
	K <sup>+</sup> (100 mM) <sup>a</sup>	hybrid +(2-tetrad) antiparallel	NMR	_____	[31]
d[AG <sub>3</sub> (T <sub>2</sub> AG <sub>3</sub> ) <sub>3</sub> ]	Na <sup>+</sup> (100 mM)	antiparallel (basket)	NMR	143D	[25]
	K <sup>+</sup> (diffusion)	parallel (propeller)	X-ray	1KF1	[26]
	Na <sup>+</sup> (145 mM)	antiparallel (basket)	CD	_____	[306]
	K <sup>+</sup> (variable) <sup>e</sup>	antiparallel (basket)	CD	_____	[306]
d[T <sub>2</sub> G <sub>3</sub> (T <sub>2</sub> AG <sub>3</sub> ) <sub>3</sub> A]	K <sup>+</sup> (100 mM) <sup>a</sup>	hybrid-1	NMR	2GKU	[27]
d[A <sub>3</sub> G <sub>3</sub> (T <sub>2</sub> AG <sub>3</sub> ) <sub>3</sub> A <sub>2</sub> ]	K <sup>+</sup> (100 mM) <sup>a</sup>	hybrid-1	NMR	2HY9	[28]
d[TAG <sub>3</sub> (T <sub>2</sub> AG <sub>3</sub> ) <sub>3</sub> ]	K <sup>+</sup> (100 mM) <sup>a</sup>	hybrid-1	NMR	2JSM	[29]
	K <sup>+</sup> (150 mM)	hybrid-1	NMR	_____	[171]
	Na <sup>+</sup> (150 mM)	hybrid-1	NMR	_____	[171]
d[TAG <sub>3</sub> (T <sub>2</sub> AG <sub>3</sub> ) <sub>3</sub> T]		hybrid-1 + hybrid-2	NMR	_____	[29]

d[TAG <sub>3</sub> (T <sub>2</sub> AG <sub>3</sub> ) <sub>3</sub> T <sub>2</sub> ]	K <sup>+</sup> (100 mM) <sup>a</sup>	hybrid-2	NMR	2JSL	[29]
d[(T <sub>2</sub> AG <sub>3</sub> ) <sub>4</sub> T <sub>2</sub> ]	K <sup>+</sup> (110 mM) <sup>f</sup>	hybrid-2	NMR	2JPZ	[30]
d[(T <sub>2</sub> AG <sub>3</sub> ) <sub>n</sub> ] <sup>g</sup>	Na <sup>+</sup> (100 mM)	antiparallel	UV-melting	—	[271]
	K <sup>+</sup> (100 mM)	parallel + antiparallel	UV-melting	—	[271]
d[G <sub>3</sub> (T <sub>2</sub> AG <sub>3</sub> ) <sub>3</sub> T] <sup>h</sup>	K <sup>+</sup> (100 mM) <sup>a</sup>	antiparallel (2-tetrad basket) <sup>i</sup>	NMR	2KF8	[31]
d[A(G <sub>3</sub> T <sub>2</sub> A) <sub>2</sub> IG <sub>2</sub> T <sub>2</sub> AG <sub>3</sub> T]	K <sup>+</sup> (110 mM) <sup>e</sup>	antiparallel (2-tetrad basket)	NMR	2KKA	[32]
r[(UUAGGG) <sub>4</sub> ]	Na <sup>+</sup> (100 mM)	parallel	CD	—	[173]
	K <sup>+</sup> (100 mM) <sup>a</sup>	parallel	CD	—	[174]

<sup>a</sup> Calculated for 70 mM KCl + 20 mM potassium phosphate (pH 7.0)

<sup>b</sup> Calculated for 200 mM NaCl + 10 mM sodium phosphate (pH 6.8)

<sup>c</sup> Calculated for 100 mM NaCl + 10 mM sodium phosphate (pH 6.8)

<sup>d</sup> Unpublished

<sup>e</sup> Addition of up to 100 mM in a 145 mM Na<sup>+</sup> solution

<sup>f</sup> Calculated for 70 mM KCl + 25 mM potassium phosphate (pH 7.0)

<sup>g</sup>  $n = 4-12$

<sup>h</sup> Other sequences adopting the same fold: G<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>, G<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>T<sub>2</sub>, and G<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>T<sub>2</sub>A

<sup>i</sup> Predominant fold (60%). Mutation of G7 to a B<sup>r</sup>G favors this topology (90%)

### 4.1.3. Summary

Clearly, K<sup>+</sup> allows d[AG<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>] to adopt a variety of structures, whereas only the basket type G-quadruplex has been solved in Na<sup>+</sup> conditions. The precise topologies adopted in potassium conditions remains controversial, the consensus being that it is likely a mixture. Other four-repeat sequences spark similar debate, notably the minimal 21-mer sequence. A current issue is that a number of studies performed prior to the publication of the antiparallel 2-tetrad structures 2KF8 and 2KKA (in 2009-2010) often attributed antiparallel signatures to the 3-tetrad antiparallel ‘basket’ topology (‘Na<sup>+</sup> fold’). In the same vein, folding/unfolding intermediates are not usually considered (see section 3.3). Shorter two-repeat sequences also exhibit a clear cation-dependent folding topology, but not the three-repeat one. To complicate the picture, the exact sequence used clearly impacts the structure(s) formed, and the analytical methods sometimes influence those detectable. Progress is needed to separate conformation mixtures while preserving each of them, in order to disentangle the cation effects. In contrast with DNA G-quadruplexes, the topologies formed by RNA sequences do not seem to exhibit any cation-dependence. Finally, most studies on long DNA or RNA telomeric sequences (reviewed in reference [172]) only involve K<sup>+</sup>, and further work is required to bring insight into the cation-dependency of G-quadruplex folding.

## 4.2 Other Sequences

Following the work of Sen and Gilbert who first investigated the cation-dependence on the topology of various sequences [315], Hardin *et al.* have shown that the telomeric sequence of *Tetrahymena* d[(T<sub>2</sub>G<sub>4</sub>)<sub>4</sub>] folds into G-quadruplexes of different structures in relatively low K<sup>+</sup> and Na<sup>+</sup> concentrations (20 mM), by a combination of NMR, CD, gel electrophoresis and an early use of SE-HPLC for G4 nucleic acids [215]. The latter method suggested that the K<sup>+</sup>-form is twice as long as the Na<sup>+</sup> form, and this was ascribed to the formation of a tetramolecular structure in K<sup>+</sup>. The intramolecular Na<sup>+</sup>-structure was solved in 1994 by Wang and Patel (PDB ID 186D) [46]. Despite its tracts of four guanines, it displays only three tetrads, linked by successive lateral-lateral-double chain reversal loops, in a pattern similar to the hybrid-2 G-quadruplex formed by the human telomeric sequence [29,30]. The 2-repeat sequence d(TG<sub>4</sub>T)<sub>2</sub> was also examined in Na<sup>+</sup>, and adopts two distinct, interconverting, bimolecular, antiparallel 4-tetrad structures, which differ by the arrangement of the lateral loops (head-to-head or head-to-tail) [176].

The one-repeat *Oxytricha nova* telomeric DNA sequence d(T<sub>4</sub>G<sub>4</sub>) forms a parallel, tetramolecular G-quadruplex in Na<sup>+</sup> or K<sup>+</sup> conditions, but oligomerizes only in the latter case [119]. The two-guanine tracts counterpart d(G<sub>4</sub>T<sub>4</sub>G<sub>4</sub>), sometimes refers to as Oxy-1.5, retains the same fold in the presence of K<sup>+</sup> or Na<sup>+</sup> these cations, whether it is studied by NMR [96,125,126] or X-Ray crystallography [127,128]. It is composed of four tetrads with the thymine loops connecting strands diagonally over external tetrads. The potassium and sodium forms differ by the position of the cations (Figure 5). Large chemical shifts differences between these species were evidenced suggesting that the geometry of the tetrads are somewhat different. Dingley *et al.* have evidenced by studying scalar couplings that (C2)NH<sub>2</sub>...N7 H-bonds are shorter in presence of Na<sup>+</sup> than K<sup>+</sup>, in agreement with the crystal structures (2.62 Å vs 2.78 Å), likely because Na<sup>+</sup> residing in the center of the tetrads holds the guanines more tightly than K<sup>+</sup> [107]. The similar sequence d(G<sub>3</sub>T<sub>4</sub>G<sub>3</sub>) also folds into a bimolecular antiparallel structure with diagonal loops in both alkali cations [316,317]. However, the loops are more flexible with K<sup>+</sup>, where they adopt two possible and interconverting conformations, and the K<sup>+</sup>-stabilized tetrads are slightly larger than with Na<sup>+</sup> [317]. Moreover, long incubation times after addition of K<sup>+</sup> on a Na<sup>+</sup> sample results in the likely formation of a tetramolecular G-quadruplex, while Na<sup>+</sup> alone does not facilitates this association.

Plavec *et al.* have shown that another related sequence, d(G<sub>4</sub>T<sub>4</sub>G<sub>3</sub>), folds into a Na<sup>+</sup>-containing G-quadruplex similar to the above-mentioned oligonucleotide, but adopts a mixture of structures with K<sup>+</sup> [318], whereas d(G<sub>3</sub>T<sub>4</sub>G<sub>4</sub>) and d(G<sub>4</sub>T<sub>4</sub>G<sub>4</sub>) adopt the same bimolecular fold with both cations, evidencing that small and targeted changes can have a dramatic effect on both the G-quadruplex topology and sensitivity to cation nature [319]. Two groups have independently demonstrated by NMR that the four-guanine tracts sequence d[G<sub>4</sub>(T<sub>4</sub>G<sub>4</sub>)<sub>3</sub>] ('Oxy-3.5') adopts an antiparallel topology in Na<sup>+</sup> (PDB ID: 201D and 230D) [175,320]. Upon addition of K<sup>+</sup> cations, significant changes in the spectra were monitored, without apparent change in the global topology, as with Oxy-1.5 [320]. Single molecule

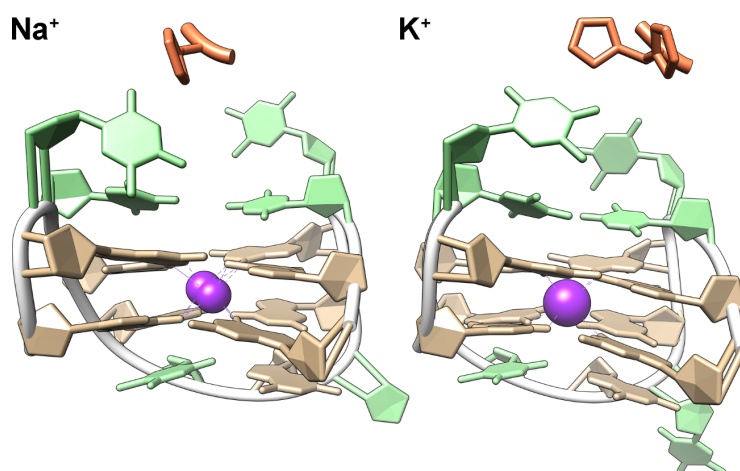
FRET spectroscopy gave consistent results with Na<sup>+</sup> that promotes an antiparallel G-quadruplex [243]. However, K<sup>+</sup> leads to a mixture of a parallel and an antiparallel structure, the antiparallel differing from the Na<sup>+</sup> fold. These different topologies were also observed with longer sequences d(T<sub>4</sub>G<sub>4</sub>)<sub>n≥4</sub> by native PAGE and UV-melting [119]. Potassium seems to lead to the formation of a mixture of parallel and antiparallel intra- and intermolecular G-quadruplexes, whereas Na<sup>+</sup> promotes the formation of antiparallel G-quadruplexes only, either intramolecular ( $n = 4\text{—}7$ ) or in a mixtures of intra- and intermolecular assemblies ( $n = 8\text{—}12$ ). In the same vein, Thomas Jr *et al.* published a phase diagram for Na<sup>+</sup> and K<sup>+</sup> of the sequence d(T<sub>4</sub>G<sub>4</sub>)<sub>4</sub> ('Oxy-4') [244]. At low concentrations, both cations promote the formation of an intramolecular antiparallel G-quadruplex. However, at higher concentrations, K<sup>+</sup> was more effective at stimulating the formation of tetramolecular assemblies. The midpoints of conversion, estimated from the Raman spectra, are 65 mM and 225 mM for K<sup>+</sup> and Na<sup>+</sup>, respectively.

More generally, a number of other telomeric sequences have distinct CD signature in Na<sup>+</sup> vs K<sup>+</sup> solutions, including *Arabidopsis*, *L. esculentum*, *C. glabrata*, *S. cerevisiae*, and *Paramecium* [114]. These sequences have G<sub>3</sub> or G<sub>4</sub> repeats, and give antiparallel signature in Na<sup>+</sup>, but not in K<sup>+</sup> conditions, where hybrid structures or mixtures are likely. Conversely, *Bombyx* and *Ascaris* have an antiparallel signature in Na<sup>+</sup> and K<sup>+</sup>. Both have repeats of two guanines only, and therefore contain most likely two tetrads, which was confirmed by NMR for *Bombyx mori* in K<sup>+</sup> conditions [68]. Guédin *et al.* examined eighty sequences containing four tracts of three guanines with loops of variable length (between 1 and 15 bases), following the template G<sub>3</sub>L<sub>a</sub>G<sub>3</sub>L<sub>b</sub>G<sub>3</sub>L<sub>c</sub>G<sub>3</sub>, where L<sub>a</sub>, L<sub>b</sub>, and L<sub>c</sub> are thymine or TTA loops, in presence of 100 mM K<sup>+</sup> or Na<sup>+</sup> [115]. In K<sup>+</sup>, all short looped sequences (two one-nucleotide loops) adopt a parallel topology, while longer loops (notably with two three-nucleotide loops) seem to promote hybrid structures, which is reminiscent of the human telomeric sequence. Even in the presence of long loops (up to nine nucleotides) that allow a certain flexibility, none of the sequences display a clear antiparallel signature. Conversely, some G-quadruplexes have an antiparallel signature in the presence of Na<sup>+</sup> (d(G<sub>3</sub>T<sub>2</sub>G<sub>3</sub>T<sub>3</sub>G<sub>3</sub>T<sub>3</sub>G<sub>3</sub>), d(G<sub>3</sub>T<sub>1</sub>G<sub>3</sub>T<sub>3</sub>G<sub>3</sub>T<sub>3</sub>G<sub>3</sub>), d(G<sub>3</sub>T<sub>4</sub>G<sub>3</sub>T<sub>4</sub>G<sub>3</sub>T<sub>4</sub>G<sub>3</sub>)), although short loops also promote hybrid folds.

Recently, Hartig and coworkers have shown by CD and NMR that the sequence d[(G<sub>4</sub>CT)<sub>3</sub>G<sub>4</sub>], from the human pathogen *Treponema pallidum*, exhibits a remarkable cation dependency [222]. Low K<sup>+</sup> concentrations promotes a 4-tetrad intramolecular antiparallel structure, while Na<sup>+</sup> and Li<sup>+</sup> do not markedly induce G<sub>4</sub> folding, even at high concentrations. The corresponding G<sub>5</sub> motif also adopts an antiparallel signatures, but not the G<sub>3</sub> counterpart that favors a parallel conformation, as does the absence of thymine in the loops. Moreover, CD, AUC and EPR experiments demonstrate that higher K<sup>+</sup> concentrations ( $\geq 500$  mM) lead to the formation of a parallel tetramer, which was confirmed by SE-HPLC shortly after [37]. Interestingly, low concentrations of K<sup>+</sup> supplemented by Na<sup>+</sup> or Li<sup>+</sup> to increase the ionic strength is sufficient to trigger the formation of the tetramer.

Russo Krauss and coworkers have obtained high resolution crystal structures of TBA (d(G<sub>2</sub>T<sub>2</sub>G<sub>2</sub>TGTG<sub>2</sub>T<sub>2</sub>G<sub>2</sub>)) bound to  $\alpha$ -thrombin in presence of Na<sup>+</sup> or K<sup>+</sup> (PDB ID: 4DIH and 4DII) (Figure 8) [161]. The difference in G-quadruplex stability in favor of the latter cation is

well known [89,98,111], as is the antiparallel ‘chair-type’ structure observed with both alkali cations. However, this work highlighted subtle cation-induced differences in the structure that may explain differences in binding mode and potency, and as a possible consequence the enhanced clotting inhibitory activity of the aptamer in presence of potassium.  $K^+$  is coordinated at the center of the cavity between the two G-tetrads and bridges together all the eight guanine O6 atoms in a distorted anti-prism geometry ( $d_{O-K} = 2.7\text{—}2.9\text{ \AA}$ ), thereby increasing the rigidity and the stability of the G-quadruplex. On the other hand,  $Na^+$  can occupy two alternative coordination sites, closer to one of the two tetrads ( $d_{O-K} = 2.2\text{—}2.7\text{ \AA}$ ), therefore conferring a higher plasticity to the aptamer that allows a better fit with the binding surface of  $\alpha$ -thrombin. Accordingly, the  $\Delta T_m$  between  $K^+$ - and  $Na^+$ -coordinating TBA decreases from 29 (free) to 12 °C (bound). Another consequence is the absence of interaction of  $TBA \cdot Na^+$  with the His71 residue, crucial for the inhibition of the fibrinogen conversion to fibrin by  $\alpha$ -thrombin, while this interaction is visible for  $TBA \cdot K^+$ .



**Figure 8** Crystal structure of TBA coordinating  $Na^+$  (left; purple spheres: two positions; PDB ID: 4DIH) or  $K^+$  (right; purple sphere: one position; PDB ID: 4DII), bound to  $\alpha$ -thrombin (not shown except for His71, in orange) [161]. Guanosines are depicted in brown, thymidines in green, and the phosphate backbones as white ribbons.

### 4.3 General trends

Very clearly, the nature of the cation impacts the structure of DNA G-quadruplexes, in terms of topology, loop geometry, and strand stoichiometry. This does not necessarily translate into a different topology, and has sometimes more subtle effects (*e.g.* cation location, loops flexibility, quartet size), but even these small modifications can lead to significant outcomes [161]. Other factors influence the structure, mainly the sequence (*i.e.* number of guanines, of repeats, length and composition of loops), the strand concentration (oligomerization), and the presence of co-solvents. In particular, small sequence alteration can dramatically change the structure(s) formed and, more importantly within the scope of this chapter, the cation-dependency of G-quadruplexes (*e.g.*  $d(G_4T_4G_3)$  vs.  $d(G_3T_4G_4)$  and  $d(G_4T_4G_4)$  [318]).

From all the studies performed on a wide range of sequences, it seems that potassium does not promote the formation of intramolecular antiparallel structures of four-repeat G<sub>3</sub> motif sequences, whereas sodium can. Conversely, four-repeat 2-tetrad G-quadruplexes seem to adopt preferentially antiparallel fold in presence of K<sup>+</sup> (e.g. human telomeres [31,32], TBA [24], *Bombyx mori* and *Ascaris* telomeres [68,114], HIV-PRO1 [69], 21CTA [45]). Sequences with G<sub>4</sub> and longer motifs are also able to adopt antiparallel conformations in K<sup>+</sup> conditions. Moreover, K<sup>+</sup> promotes more efficiently oligomerization and the formation of tetramolecular assemblies than Na<sup>+</sup> [39,222], although (i) oligomerization in Na<sup>+</sup> solutions has been observed [245], and (ii) it should be noted that a number of observations have been made with high-strand-concentration samples, and may not be reflected in dilute conditions. Finally, G-rich RNA sequences fold in parallel G-quadruplexes regardless of the cation, although differences have been noticed regarding their propensity to oligomerize.

## 5 Cation-dependent conformational switching

The first example of cation-dependent structural switch of G-quadruplexes was reported in 1990 by Sen and Gilbert [315]. Although the authors based their conclusions on native PAGE experiments only, and discarded intramolecular G-quadruplexes as intermediates structures or by-products, it clearly shows that sodium and potassium can dramatically affect the structure(s) adopted by guanine-rich oligonucleotides, as well as their rate of formation. In 1992, Hardin *et al.* provided a nice example of cation-dependent hairpin-to-quadruplex conversion, which led to ranking of cations as a function of their G-quadruplex stabilization properties: **K<sup>+</sup> > Ca<sup>2+</sup> > Na<sup>+</sup> > Mg<sup>2+</sup> > Li<sup>+</sup>**, and **K<sup>+</sup> > Rb<sup>+</sup> > Cs<sup>+</sup>** [321].

We have seen in sections 3 and 4 that the nature and concentration of the cation(s) has a large effect on the stability and structure of G-quadruplex-forming sequences. It is therefore not surprising to find a number of examples of cation-triggered conformational switch from a less stable structure coordinating a cation to a more stable structure coordinating another cation. Typically, the switch is triggered by adding a more stabilizing cation, hence often by adding K<sup>+</sup> to a Na<sup>+</sup>-containing solution.

The telomeric sequence is well known for its particularly pronounced polymorphism in presence of Na<sup>+</sup> or K<sup>+</sup> (section 4.1), so the conversion from the sodium to the potassium form(s) is extensively studied. In 2006, Yang *et al.* proposed that the conversion from the Na<sup>+</sup>-basket form to the K<sup>+</sup>-hybrid form of human telomeric sequences involves the cation exchange, followed by a partial unfolding and restructuration [165]. Partial unfolding possibly leads to triplex intermediates, as also suggested by Sugiyama *et al.* and Chaires *et al.* [282,285-287]. Gray *et al.* have shown that the (30 mM) Na<sup>+</sup> to (50 mM) K<sup>+</sup> exchange of the 22-mer d[AG<sub>3</sub>(G<sub>3</sub>T<sub>2</sub>A)<sub>3</sub>] is followed by a conformation change, from the Na<sup>+</sup>-basket type to the K<sup>+</sup>-hybrid (Figure 6C) [288]. Three distinct kinetic processes are involved: (i) A fast cation exchange ( $\tau_1 \sim 250 \mu\text{s}$ ) yields a K<sup>+</sup>-coordinating basket G-quadruplex 'I1', (ii) A partial unfolding giving two possible triplex intermediates 'I2' (opening in 5' or 3';  $\tau_2 \sim 50 \text{ s}$ ), and (iii) a slower ( $\tau_3 \sim 800 \text{ s}$ ) refolding that converts a diagonal loop to a lateral loop, yielding

hybrid-1 and hybrid-2 K<sup>+</sup>-stabilized G-quadruplexes (opening in 5' gives the hybrid-1, 3' the hybrid-2). The free energy barrier between the starting and final structures is relatively modest (1.4—2.4 kcal.mol<sup>-1</sup>). These results were reassessed recently under the light of new folding pathway data, with  $\tau_2 > \tau_3$  [285]. The presence of a folding intermediate (presumably a triplex) was also suggested in Na<sup>+</sup> solution for the 21-mer sequence d[G<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>] [322].

Worth mentioning, Chang *et al.* have argued that the Na<sup>+</sup> to K<sup>+</sup> exchange of the 23-mer sequence d[TAG<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>] does not involve a triplex intermediate, nor any unfolding, since the sequence adopts the same hybrid-1 topology in both K<sup>+</sup> and Na<sup>+</sup> solutions (Figure 6B) [171]. They postulate that either slightly different human telomeric sequences adopt different exchange mechanism or that other studies have not satisfyingly characterized the initial and final states. Besides cation nature, a change in cation concentration can also trigger structural changes. Hartig *et al.* have shown that the sequence d[(G<sub>4</sub>CT)<sub>3</sub>G<sub>4</sub>], from the human pathogen *Treponema pallidum* promotes a 4-tetrad intramolecular antiparallel structure at low K<sup>+</sup> concentrations, while Na<sup>+</sup> and Li<sup>+</sup> does not markedly induce G4 folding, even at high concentrations [222]. Interestingly, CD, AUC and EPR experiments demonstrated that higher K<sup>+</sup> concentrations ( $\geq 500$  mM) lead to the formation of a parallel tetramolecular assembly, which was also observed by SE-HPLC in an independent study [37]. Interestingly, low concentrations of K<sup>+</sup> supplemented by Na<sup>+</sup> or Li<sup>+</sup> to increase the ionic strength is sufficient to trigger the formation of the tetramer, possibly to screen the negatively charged phosphate repulsions upon strand association although the authors did not comment on that.

Abu-Ghazalah and coworkers have observed the structural conversion of human telomeric sequences d[G<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>] ('basket'), d[A<sub>3</sub>G<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>A<sub>2</sub>] ('hybrid-1'), and d[T<sub>2</sub>AG<sub>3</sub>(T<sub>2</sub>AG<sub>3</sub>)<sub>3</sub>T<sub>2</sub>] ('hybrid-2') at high strand concentrations (2 mM) [245]. CD suggested the formation of parallel tetramolecular aggregates for the former sequence in 100 mM Na<sup>+</sup> solutions, with a relaxation time of around 10 hours, which can re-dissociate rapidly upon dilution. The conversion takes place regardless of the Na<sup>+</sup> concentration, however the rate of conversion increases with increasing strand and Na<sup>+</sup> concentrations. The two latter hybrid-type sequences can also aggregate, albeit only at high salt concentrations (1 M K<sup>+</sup>).

## 6 Concluding remarks and future directions

G-quadruplexes' relevance to life sciences was revealed in the late 1980s with seminal works on telomeric sequences [81,82,315,323]. The role of monovalent cations was reviewed as early as 1991 [301], but an increasing number of other sequences have been investigated since. Bioinformatics studies suggest the presence of a very large number of putative G-quadruplex-forming sequences in the human genome only, ranging from 370 000 [4-6] to more than a million (Bedrat *et al.*, unpublished results). Compared to this very large sequence space, only a few sequences have been explored in depth, but several (six) very different topologies have been observed thus far [22,23]. Many studies have dealt with the influence of alkali cation on G-quadruplex structures, as can be seen from the large range of publications cited in this chapter. A marked difference of stability between K<sup>+</sup>, Na<sup>+</sup> and Li<sup>+</sup> is indeed

widely accepted as a good indication that G-quadruplex structures are involved.

Quadruplexes are often considered to operate as allosteric switches, implicated notably in gene regulation processes [7-15]. Hence, the interaction of intracellular potassium and sodium with G-rich sequences might play an important role in the regulation of biological processes. As a result, the importance of alkali metal coordination, most notably of sodium and potassium, on the stability and topology of structures formed by G-quadruplex-prone sequences, was intensively investigated, but there is a clear bias in favor of potassium because of their prevalence in cells. These studies were also prompted for other purposes such as artificial switch elements for DNA-based nanodevices, but to a far lower extent.

Comparison of  $\text{Na}^+$  and  $\text{K}^+$  in terms of G-quadruplex stabilization and structures yielded fairly consistent results. Potassium stabilizes G-quadruplexes better than sodium and the reasons are well understood. Cation coordination is the driving force towards G-quadruplex folding or interconversion, in accordance with the unfolded state of guanine-rich sequence in absence of suitable cation [62,285]. However, the question as to why a sequence folds into one or several structures in particular, in presence of a given cation, remains largely unanswered. Answering this question is difficult because the structure depends not only on the nature of the cation, but also on cation concentration, strand concentration, and temperature. Some basic trends are very well known: tetramolecular G-quadruplexes are parallel, RNA G-quadruplexes are parallel, regardless of the cation nature. However, when it comes to the folding of intramolecular DNA G-quadruplexes, the most interesting when it comes to studying genomes, venturing an educated guess seems very risky. In fact, even the structures of some heavily studied sequences such as the human telomeric sequence in potassium solutions are still controversial. Numerous studies clearly show that minor sequence alterations lead to large structural effects, and the human telomeric sequence is the epitome of this phenomenon [172,203].

Similarly, the pathways of folding/unfolding and interconversion of G-quadruplexes are still a matter of debate. No clear view of cation effects emerged yet, except that potassium can swiftly displace sodium, whether it is accompanied by structural alterations or not. G-quadruplexes are often presented as biological switches within the genome because of their ability to be either folded or unfolded (and in the latter case, hybridized in a canonical duplex). Moreover, even small differences in the structure of G-quadruplexes due to the binding of  $\text{K}^+$  or  $\text{Na}^+$  can have an effect on protein binding [161]. In this context, studies mimicking the cellular environment are needed (crowding agents, proteins, complex mixtures of cations), complemented by the use of appropriate oligonucleotides (long telomeric sequences, G-quadruplexes embedded in duplex matrices). Recent studies have started to tackle these issues [169,242,314,324], but the complexity of these experiments makes it difficult to assess the influence of the cations in cell-like environments. The elucidation alkali cation effects on G-quadruplex nucleic acids structure and stability is a difficult task, yet an important milestone towards the prediction of structure from the sequence and environment, and towards the design of stimuli-responsive artificial DNA switches.



## Abbreviations and Definitions

AUC	analytical ultracentrifugation
CD	circular dichroism
DFT-D	dispersion-corrected density functional theory
DNA	deoxyribonucleic acid
DOSY	diffusion ordered spectroscopy
DSC	differential scanning calorimetry
EPR	electron paramagnetic resonance
ESI-MS	electrospray ionization mass spectrometry
FRET	Förster resonance energy transfer
G4	quadruplex nucleic acid
GMP	guanosine 5'-monophosphate
HSQC	heteronuclear single-quantum correlation spectroscopy
IDS	isothermal difference spectra
IMS-MS	ion-mobility spectrometry mass spectrometry
ITC	isothermal titration calorimetry
LNA	locked nucleic acid
MALDI-TOF	matrix-assisted laser desorption/ionization time-of-flight mass spectrometry
MAS	magic-angle spinning
MD	molecular dynamics
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
nt	nucleotide
PAGE	polyacrylamide gel electrophoresis
PBFI	benzofuran-isophthalate crown ether indicator
PDB	protein data bank
PNA	peptide nucleic acid
QM	quantum mechanics
RMSD	root-mean-square deviation
RNA	ribonucleic acid
SE-HPLC	size-exclusion high-performance liquid chromatography
STM	scanning tunneling microscope
TBA	thrombin binding aptamer
TDS	thermal difference spectra
TERRA	telomeric repeat-containing RNA
TMAA	trimethyl ammonium acetate
UTR	untranslated region

**Acknowledgment:** Pictures of G-quadruplexes from the PDB entries were generated with UCSF Chimera (alpha version 1.11) [325]. Funding was provided by *Agence Nationale de la*

*Recherche* (OligoSwitch [ANR- 12-IS07–0001], ‘Quarpediem’ [ANR-12-BSV8–0008–01], and ‘VIBBnano’ [ANR-10-NANO-04–03]).

## References

1. J. T. Davis, *Angew. Chem. Int. Ed. Engl.* **2004**, *43*, 668-698.
2. *Quadruplex Nucleic Acids*, Eds. S. Neidle, S. Balasubramanian, Royal Society of Chemistry, Cambridge, UK, 2006.
3. *Quadruplex Nucleic Acids*, Eds. J. B. Chaires, D. Graves, Springer, Berlin, 2013.
4. J. L. Huppert, S. Balasubramanian, *Nucleic Acids Res.* **2005**, *33*, 2908-2916.
5. J. L. Huppert, S. Balasubramanian, *Nucleic Acids Res.* **2007**, *35*, 406-413.
6. A. K. Todd, *Nucleic Acids Res.* **2005**, *33*, 2901-2907.
7. G. Biffi, M. Di Antonio, D. Tannahill, S. Balasubramanian, *Nature Chem.* **2013**, *6*, 75-80.
8. G. Biffi, D. Tannahill, J. McCafferty, S. Balasubramanian, *Nature Chem.* **2013**, *5*, 182-186.
9. E. Y. N. Lam, D. Beraldi, D. Tannahill, S. Balasubramanian, *Nat. Commun.* **2013**, *4*, 1796-1796.
10. J.-L. Mergny, *Nat. Chem. Biol.* **2012**, *8*, 225-226.
11. S. Millevoi, H. Moine, S. Vagner, *Wiley Interdiscip. Rev.: RNA* **2012**, *3*, 495-507.
12. H. J. Lipps, D. Rhodes, *Trends Cell Biol.* **2009**, *19*, 414-422.
13. L. T. Gray, A. C. Vallur, J. Eddy, N. Maizels, *Nat. Chem. Biol.* **2014**, *10*, 313-318.
14. A. Baral, P. Kumar, R. Pathak, S. Chowdhury, *Mol. BioSyst.* **2013**, *9*, 1568-1575.
15. M. Metifiot, S. Amrane, S. Litvak, M. L. Andreola, *Nucleic Acids Res.* **2014**, *42*, 12352-12366.
16. S. Balasubramanian, S. Neidle, *Curr. Opin. Chem. Biol.* **2009**, *13*, 345-353.
17. T. M. Ou, Y. J. Lu, J. H. Tan, Z. S. Huang, K. Y. Wong, L. Q. Gu, *ChemMedChem* **2008**, *3*, 690-713.
18. S. A. Ohnmacht, S. Neidle, *Bioorg. Med. Chem. Lett.* **2014**, *24*, 2602-2612.
19. G. N. Parkinson, in *Guanine Quartets: Structure and Application*, Eds. L. Spindler, W. Fritzsche, Royal Society of Chemistry, Cambridge, UK, 2012, pp. 237-247.
20. D. J. Patel, A. T. Phan, V. Kuryavyi, *Nucleic Acids Res.* **2007**, *35*, 7429-7455.
21. > 8000 publications including the word quadruplex or tetraplex in their title, as of January 2015, according to Thomson Reuters' Web of Science, with more than 700 for 2014 only.
22. M. Webba da Silva, *Chem. --Eur. J.* **2007**, *13*, 9738-9745.
23. A. I. Karsisiotis, C. O'Kane, M. Webba da Silva, *Methods* **2013**, *64*, 28-35.
24. P. Schultze, R. F. Macaya, J. Feigon, *J. Mol. Biol.* **1994**, *235*, 1532-1547.
25. Y. Wang, D. J. Patel, *Structure* **1993**, *1*, 263-282.
26. G. N. Parkinson, M. P. H. Lee, S. Neidle, *Nature* **2002**, *417*, 876-880.

27. K. N. Luu, A. T. Phan, V. Kuryavyi, L. Lacroix, D. J. Patel, *J. Am. Chem. Soc.* **2006**, *128*, 9963-9970.
28. J. Dai, C. Punchihewa, A. Ambrus, D. Chen, R. A. Jones, D. Yang, *Nucleic Acids Res.* **2007**, *35*, 2440-2450.
29. A. T. Phan, V. Kuryavyi, K. N. Luu, D. J. Patel, *Nucleic Acids Res.* **2007**, *35*, 6517-6525.
30. J. Dai, M. Carver, C. Punchihewa, R. A. Jones, D. Yang, *Nucleic Acids Res.* **2007**, *35*, 4927-4940.
31. 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.
32. Z. Zhang, J. Dai, E. Veliath, R. A. Jones, D. Yang, *Nucleic Acids Res.* **2010**, *38*, 1009-1021.
33. J. Zhou, K. Murayama, S. Amrane, F. Rosu, H. Kashida, A. Bourdoncle, H. Asanuma, J.-L. Mergny, *Chem. Sci.* **2013**, *4*, 3693-3693.
34. V. Esposito, A. Virgilio, A. Randazzo, A. Galeone, L. Mayol, *Chem. Commun.* **2005**, 3953-3955.
35. V. Esposito, A. Virgilio, A. Pepe, G. Oliviero, L. Mayol, A. Galeone, *Bioorg. Med. Chem.* **2009**, *17*, 1997-2001.
36. M. Webba da Silva, *Methods* **2007**, *43*, 264-277.
37. E. Largy, J. L. Mergny, *Nucleic Acids Res.* **2014**, *42*, e149.
38. P. Tothova, P. Krafcikova, V. Viglasky, *Biochemistry* **2014**, *53*, 7013-7027.
39. N. Smargiasso, F. Rosu, W. Hsia, P. Colson, E. S. Baker, M. T. Bowers, E. De Pauw, V. Gabelica, *J. Am. Chem. Soc.* **2008**, *130*, 10208-10216.
40. J. Zhou, A. Bourdoncle, F. Rosu, V. Gabelica, J.-L. Mergny, *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 11002-11005.
41. V. T. Mukundan, A. T. Phan, *J. Am. Chem. Soc.* **2013**, *135*, 5017-5028.
42. H. Martadinata, A. T. Phan, *Biochemistry* **2014**, *53*, 1595-1600.
43. M. Adrian, D. J. Ang, C. J. Lech, B. Heddi, A. Nicolas, A. T. Phan, *J. Am. Chem. Soc.* **2014**, *136*, 6297-6305.
44. G. D. Balkwill, T. P. Garner, H. E. Williams, M. S. Searle, *J. Mol. Biol.* **2009**, *385*, 1600-1615.
45. K. W. Lim, P. Alberti, A. Guédin, L. Lacroix, J.-F. Riou, N. J. Royle, J.-L. Mergny, A. T. Phan, *Nucleic Acids Res.* **2009**, *37*, 6239-6248.
46. Y. Wang, D. J. Patel, *Structure* **1994**, *2*, 1141-1156.
47. N. Zhang, A. Gorin, A. Majumdar, A. Kettani, N. Chernichenko, E. Skripkin, D. J. Patel, *J. Mol. Biol.* **2001**, *311*, 1063-1079.
48. A. Kettani, A. Gorin, A. Majumdar, T. Hermann, E. Skripkin, H. Zhao, R. Jones, D. J. Patel, *J. Mol. Biol.* **2000**, *297*, 627-644.
49. T. Mashima, A. Matsugami, F. Nishikawa, S. Nishikawa, M. Katahira, *Nucleic Acids Res.* **2009**, *37*, 6249-6258.
50. H. Martadinata, A. T. Phan, *Biochemistry* **2013**, *52*, 2176-2183.
51. A. Matsugami, K. Ouhashi, M. Kanagawa, H. Liu, S. Kanagawa, S. Uesugi, M.

- Katahira, *J. Mol. Biol.* **2001**, *313*, 255-269.
52. N. Borbone, J. Amato, G. Oliviero, V. D'Atri, V. Gabelica, E. De Pauw, G. Piccialli, L. Mayol, *Nucleic Acids Res.* **2011**, *39*, 7848-7857.
  53. A. T. Phan, V. Kuryavyi, S. Burge, S. Neidle, D. J. Patel, *J. Am. Chem. Soc.* **2007**, *129*, 4386-4392.
  54. D. J. Patel, A. T. Phan, V. Kuryavyi, *Nucleic Acids Res.* **2007**, *35*, 7429-7455.
  55. D. E. Gilbert, J. Feigon, *Curr. Opin. Struct. Biol.* **1999**, *9*, 305-314.
  56. S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd, S. Neidle, *Nucleic Acids Res.* **2006**, *34*, 5402-5415.
  57. J. L. Huppert, *Philos. Trans. R. Soc., A* **2007**, *365*, 2969-2984.
  58. J. L. Huppert, *Chem. Soc. Rev.* **2008**, *37*, 1375-1384.
  59. *Guanine Quartets: Structure and Application*, Eds. W. Fritzsche, L. Spindler, Royal Society of Chemistry, Cambridge, 2012.
  60. P. Alberti, A. Bourdoncle, B. Saccà, L. Lacroix, J.-L. Mergny, *Org. Biomol. Chem.* **2006**, *4*, 3383-3391.
  61. C. M. Olsen, W. H. Gmeiner, L. A. Marky, *J. Phys. Chem. B* **2006**, *110*, 6962-6969.
  62. A. N. Lane, J. B. Chaires, R. D. Gray, J. O. Trent, *Nucleic Acids Res.* **2008**, *36*, 5482-5515.
  63. X. Cang, J. Šponer, T. E. Cheatham, *Nucleic Acids Res.* **2011**, *39*, 4499-4512.
  64. J. Šponer, A. Mládek, N. Spačková, X. Cang, T. E. Cheatham, S. Grimme, *J. Am. Chem. Soc.* **2013**, *135*, 9785-9796.
  65. C. J. Lech, B. Heddi, A. T. Phan, *Nucleic Acids Res.* **2013**, *41*, 2034-2046.
  66. D. Zhang, T. Huang, P. S. Lukeman, P. J. Paukstelis, *Nucleic Acids Res.* **2014**, *42*, 13422-13429.
  67. G. N. Parkinson, R. Ghosh, S. Neidle, *Biochemistry* **2007**, *46*, 2390-2397.
  68. S. Amrane, R. W. L. Ang, Z. M. Tan, C. Li, J. K. C. Lim, J. M. W. Lim, K. W. Lim, A. T. Phan, *Nucleic Acids Res.* **2009**, *37*, 931-938.
  69. S. Amrane, A. Kerkour, A. Bedrat, B. Vialet, M.-L. Andréola, J.-L. Mergny, *J. Am. Chem. Soc.* **2014**, *136*, 5249-5252.
  70. V. Kuryavyi, A. T. Phan, D. J. Patel, *Nucleic Acids Res.* **2010**, *38*, 6757-6773.
  71. R. Otero, M. Schöck, L. M. Molina, E. Laegsgaard, I. Stensgaard, B. Hammer, F. Besenbacher, *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 2270-2275.
  72. C. Fonseca Guerra, H. Zijlstra, G. Paragi, F. M. Bickelhaupt, *Chem. --Eur. J.* **2011**, *17*, 12612-12622.
  73. 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.
  74. H. K. Frensdorff, *J. Am. Chem. Soc.* **1971**, *93*, 600-606.
  75. A. Delville, C. Detellier, P. Laszlo, *J. Magn. Reson. (1969-1992)* **1979**, *34*, 301-315.
  76. E. Bouhoutsos-Brown, C. L. Marshall, T. J. Pinnavaia, *J. Am. Chem. Soc.* **1982**, *104*, 6576-6584.
  77. S. B. Zimmerman, G. H. Cohen, D. R. Davies, *J. Mol. Biol.* **1975**, *92*, 181-192.
  78. H. T. Miles, J. Frazier, *J. Am. Chem. Soc.* **1978**, *100*, 8037-8038.

79. F. B. Howard, H. T. Miles, *Biopolymers* **1982**, *21*, 147-157.
80. F. B. Howard, H. T. Miles, *Biochemistry* **1982**, *21*, 6736-6745.
81. E. Henderson, C. C. Hardin, S. K. Walk, I. Tinoco, E. H. Blackburn, *Cell* **1987**, *51*, 899-908.
82. J. R. Williamson, M. K. Raghuraman, T. R. Cech, *Cell* **1989**, *59*, 871-880.
83. E. A. Venczel, D. Sen, *Biochemistry* **1993**, *32*, 6220-6228.
84. B. I. Kankia, L. A. Marky, *J. Am. Chem. Soc.* **2001**, *123*, 10799-10804.
85. A. Włodarczyk, P. Grzybowski, A. Patkowski, A. Dobek, *J. Phys. Chem. B* **2005**, *109*, 3594-3605.
86. J. R. Williamson, *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 703-730.
87. W. S. Ross, C. C. Hardin, *J. Am. Chem. Soc.* **1994**, *116*, 6070-6080.
88. N. V. Hud, F. W. Smith, F. A. Anet, J. Feigon, *Biochemistry* **1996**, *35*, 15383-15390.
89. B. Saccà, L. Lacroix, J. L. Mergny, *Nucleic Acids Res.* **2005**, *33*, 1182-1192.
90. J. C. Bowman, T. K. Lenz, N. V. Hud, L. D. Williams, *Curr. Opin. Struct. Biol.* **2012**, *22*, 262-272.
91. R. D. Gray, J. B. Chaires, *Biophys. Chem.* **2011**, *159*, 205-209.
92. E. Galezowska, A. Gluszynska, B. Juskowiak, *J. Inorg. Biochem.* **2007**, *101*, 678-685.
93. W. Xu, Q. Tan, M. Yu, Q. Sun, H. Kong, E. Lægsgaard, I. Stensgaard, J. Kjems, J.-G. Wang, C. Wang, F. Besenbacher, *Chem. Commun.* **2013**, *49*, 7210-7212.
94. A. Risitano, K. R. Fox, *Biochemistry* **2003**, *42*, 6507-6513.
95. R. D. Gray, J. B. Chaires, *Nucleic Acids Res.* **2008**, *36*, 4191-4203.
96. P. Schultze, N. V. Hud, F. W. Smith, J. Feigon, *Nucleic Acids Res.* **1999**, *27*, 3018-3028.
97. N. V. Hud, P. Schultze, V. Sklenár, J. Feigon, *J. Mol. Biol.* **1999**, *285*, 233-243.
98. N. Kumar, S. Maiti, *Biochem. Biophys. Res. Commun.* **2004**, *319*, 759-767.
99. J. M. Wilcox, D. L. Rempel, M. L. Gross, *Anal. Chem.* **2008**, *80*, 2365-2371.
100. T. L. Niederhauser, B. R. Brown, S. P. Ziemer, J. D. Sargent, E. M. Woolley, *J. Chem. Thermodyn.* **2004**, *36*, 1067-1077.
101. Y. Wu, M. Tabata, *J. Solution Chem.* **2004**, *33*, 777-795.
102. S. P. Ziemer, T. L. Niederhauser, E. M. Woolley, *J. Chem. Thermodyn.* **2005**, *37*, 1071-1084.
103. A. Risitano, K. R. Fox, *Bioorg. Med. Chem. Lett* **2005**, *15*, 2047-2050.
104. J. A. Walmsley, T. J. Pinnavaia, *Biophys. J.* **1982**, *38*, 315-318.
105. J. Gu, J. Leszczynski, *J. Phys. Chem. A* **2002**, *106*, 529-532.
106. V. K. Misra, D. E. Draper, *Biopolymers* **1998**, *48*, 113-135.
107. A. J. Dingley, R. D. Peterson, S. Grzesiek, J. Feigon, *J. Am. Chem. Soc.* **2005**, *127*, 14466-14472.
108. M. Meyer, A. Hocquet, J. Sühnel, *J. Comput. Chem.* **2005**, *26*, 352-364.
109. P. Balagurumoorthy, S. K. Brahmachari, D. Mohanty, M. Bansal, V. Sasisekharan, *Nucleic Acids Res.* **1992**, *20*, 4061-4067.
110. P. Balagurumoorthy, S. K. Brahmachari, *J. Biol. Chem.* **1994**, *269*, 21858-21869.
111. J. L. Mergny, A. T. Phan, L. Lacroix, *FEBS Lett.* **1998**, *435*, 74-78.

112. L. Petraccone, E. Erra, V. Esposito, A. Randazzo, L. Mayol, L. Nasti, G. Barone, C. Giancola, *Biochemistry* **2004**, *43*, 4877-4884.
113. J. L. Mergny, J. Li, L. Lacroix, S. Amrane, J. B. Chaires, *Nucleic Acids Res.* **2005**, *33*, e138.
114. P. L. T. Tran, J. L. Mergny, P. Alberti, *Nucleic Acids Res.* **2011**, *39*, 3282-3294.
115. A. Guédin, J. Gros, P. Alberti, J. L. Mergny, *Nucleic Acids Res.* **2010**, *38*, 7858-7868.
116. A. Guédin, P. Alberti, J. L. Mergny, *Nucleic Acids Res.* **2009**, *37*, 5559-5567.
117. J. L. Mergny, A. De Cian, A. Ghelab, B. Saccà, L. Lacroix, *Nucleic Acids Res.* **2005**, *33*, 81-94.
118. K. W. Lim, L. Lacroix, D. J. Yue, J. K. Lim, J. M. Lim, A. T. Phan, *J. Am. Chem. Soc.* **2010**, *132*, 12331-12342.
119. M. Lu, Q. Guo, N. R. Kallenbach, *Biochemistry* **1992**, *31*, 2455-2459.
120. K. Phillips, Z. Dauter, A. I. Murchie, D. M. Lilley, B. Luisi, *J. Mol. Biol.* **1997**, *273*, 171-182.
121. D. Wei, A. K. Todd, M. Zloh, M. Gunaratnam, G. N. Parkinson, S. Neidle, *J. Am. Chem. Soc.* **2013**, *135*, 19319-19329.
122. C. Creze, B. Rinaldi, R. Haser, P. Bouvet, P. Gouet, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2007**, *63*, 682-688.
123. G. R. Clark, P. D. Pytel, C. J. Squire, S. Neidle, *J. Am. Chem. Soc.* **2003**, *125*, 4066-4067.
124. G. R. Clark, P. D. Pytel, C. J. Squire, *Nucleic Acids Res.* **2012**, *40*, 5731-5738.
125. F. W. Smith, J. Feigon, *Nature* **1992**, *356*, 164-168.
126. P. Schultze, F. W. Smith, J. Feigon, *Structure* **1994**, *2*, 221-233.
127. M. P. Horvath, S. C. Schultz, *J. Mol. Biol.* **2001**, *310*, 367-377.
128. S. Haider, G. N. Parkinson, S. Neidle, *J. Mol. Biol.* **2002**, *320*, 189-200.
129. M. L. Gill, S. A. Strobel, J. P. Loria, *Nucleic Acids Res.* **2006**, *34*, 4506-4514.
130. S. Neidle, *Methods* **2012**, *57*, 1-2.
131. M. Adrian, B. Heddi, A. T. Phan, *Methods* **2012**, *57*, 11-24.
132. S. Neidle, G. N. Parkinson, *Biochimie* **2008**, *90*, 1184-1196.
133. N. H. Campbell, G. N. Parkinson, *Methods* **2007**, *43*, 252-263.
134. J. Feigon, K. M. Koshlap, F. W. Smith, *Methods Enzymol.* **1995**, *261*, 225-255.
135. G. Laughlan, A. I. Murchie, D. G. Norman, M. H. Moore, P. C. Moody, D. M. Lilley, B. Luisi, *Science* **1994**, *265*, 520-524.
136. B. Pan, Y. Xiong, K. Shi, J. Deng, M. Sundaralingam, *Structure* **2003**, *11*, 815-823.
137. B. Pan, Y. Xiong, K. Shi, M. Sundaralingam, *Structure* **2003**, *11*, 1423-1430.
138. B. Pan, Y. Xiong, K. Shi, M. Sundaralingam, *Structure* **2003**, *11*, 825-831.
139. B. Pan, K. Shi, M. Sundaralingam, *J. Mol. Biol.* **2006**, *363*, 451-459.
140. B. Pan, K. Shi, M. Sundaralingam, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 3130-3134.
141. M. P. H. Lee, G. N. Parkinson, P. Hazel, S. Neidle, *J. Am. Chem. Soc.* **2007**, *129*, 10106-10107.
142. C. Cáceres, G. Wright, C. Gouyette, G. Parkinson, J. A. Subirana, *Nucleic Acids Res.*

- 2004**, 32, 1097-1102.
143. I. Russo Krauss, G. N. Parkinson, A. Merlino, C. A. Mattia, A. Randazzo, E. Novellino, L. Mazzarella, F. Sica, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2014**, 70, 362-370.
  144. J. Kondo, W. Adachi, S. Umeda, T. Sunami, A. Takenaka, *Nucleic Acids Res.* **2004**, 32, 2541-2549.
  145. P. K. Mandal, G. W. Collie, B. Kauffmann, I. Huc, *Angew. Chem. Int. Ed. Engl.* **2014**, 53, 14424-14427.
  146. S. M. Haider, G. N. Parkinson, S. Neidle, *J. Mol. Biol.* **2003**, 326, 117-125.
  147. P. Hazel, G. N. Parkinson, S. Neidle, *J. Am. Chem. Soc.* **2006**, 128, 5480-5487.
  148. N. H. Campbell, G. N. Parkinson, A. P. Reszka, S. Neidle, *J. Am. Chem. Soc.* **2008**, 130, 6722-6724.
  149. N. H. Campbell, M. Patel, A. B. Tofa, R. Ghosh, G. N. Parkinson, S. Neidle, *Biochemistry* **2009**, 48, 1675-1680.
  150. G. W. Collie, S. M. Haider, S. Neidle, G. N. Parkinson, *Nucleic Acids Res.* **2010**, 38, 5569-5580.
  151. G. W. Collie, S. Sparapani, G. N. Parkinson, S. Neidle, *J. Am. Chem. Soc.* **2011**, 133, 2721-2728.
  152. N. H. Campbell, D. L. Smith, A. P. Reszka, S. Neidle, D. O'Hagan, *Org. Biomol. Chem.* **2011**, 9, 1328-1331.
  153. N. H. Campbell, N. H. A. Karim, G. N. Parkinson, M. Gunaratnam, V. Petrucci, A. K. Todd, R. Vilar, S. Neidle, *J. Med. Chem.* **2012**, 55, 209-222.
  154. G. W. Collie, R. Promontorio, S. M. Hampel, M. Micco, S. Neidle, G. N. Parkinson, *J. Am. Chem. Soc.* **2012**, 134, 2723-2731.
  155. 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, 20446-20456.
  156. 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, 2959-2974.
  157. C. Bazzicalupi, M. Ferraroni, A. R. Bilia, F. Scheggi, P. Gratteri, *Nucleic Acids Res.* **2013**, 41, 632-638.
  158. G. N. Parkinson, F. Cuenca, S. Neidle, *J. Mol. Biol.* **2008**, 381, 1145-1156.
  159. D. Wei, G. N. Parkinson, A. P. Reszka, S. Neidle, *Nucleic Acids Res.* **2012**, 40, 4691-4700.
  160. D. Wei, J. Husby, S. Neidle, *Nucleic Acids Res.* **2015**, 43, 629-644.
  161. I. Russo Krauss, A. Merlino, A. Randazzo, E. Novellino, L. Mazzarella, F. Sica, *Nucleic Acids Res.* **2012**, 40, 8119-8128.
  162. S. M. Haider, S. Neidle, G. N. Parkinson, *Biochimie* **2011**, 93, 1239-1251.
  163. *Nucleic Acid-Metal Ion Interactions*, Ed. N. V. Hud, Royal Society of Chemistry, Cambridge, 2008, pp. 433.
  164. C. Bardin, J. L. Leroy, *Nucleic Acids Res.* **2008**, 36, 477-488.
  165. A. Ambrus, D. Chen, J. Dai, T. Bialis, R. A. Jones, D. Yang, *Nucleic Acids Res.* **2006**, 34, 2723-2735.
  166. A. Matsugami, Y. Xu, Y. Noguchi, H. Sugiyama, M. Katahira, *FEBS J.* **2007**, 274,

- 3545-3556.
167. A. T. Phan, K. N. Luu, D. J. Patel, *Nucleic Acids Res.* **2006**, *34*, 5715-5719.
  168. N. Zhang, A. T. Phan, D. J. Patel, *J. Am. Chem. Soc.* **2005**, *127*, 17277-17285.
  169. D. J. Yue, K. W. Lim, A. T. Phan, *J. Am. Chem. Soc.* **2011**, *133*, 11462-11465.
  170. A. T. Phan, D. J. Patel, *J. Am. Chem. Soc.* **2003**, *125*, 15021-15027.
  171. Z. F. Wang, M. H. Li, S. T. Hsu, T. C. Chang, *Nucleic Acids Res.* **2014**, *42*, 4723-4733.
  172. A. T. Phan, *FEBS J.* **2010**, *277*, 1107-1117.
  173. Y. Xu, K. Kaminaga, M. Komiyama, *J. Am. Chem. Soc.* **2008**, *130*, 11179-11184.
  174. H. Martadinata, A. T. Phan, *J. Am. Chem. Soc.* **2009**, *131*, 2570-2578.
  175. Y. Wang, D. J. Patel, *J. Mol. Biol.* **1995**, *251*, 76-94.
  176. A. T. Phan, Y. S. Modi, D. J. Patel, *J. Mol. Biol.* **2004**, *338*, 93-102.
  177. L. Hu, K. W. Lim, S. Bouaziz, A. T. Phan, *J. Am. Chem. Soc.* **2009**, *131*, 16824-16831.
  178. S. T. Hsu, P. Varnai, A. Bugaut, A. P. Reszka, S. Neidle, S. Balasubramanian, *J. Am. Chem. Soc.* **2009**, *131*, 13399-13409.
  179. A. Ambrus, D. Chen, J. Dai, R. A. Jones, D. Yang, *Biochemistry* **2005**, *44*, 2048-2058.
  180. J. Dai, D. Chen, R. A. Jones, L. H. Hurley, D. Yang, *Nucleic Acids Res.* **2006**, *34*, 5133-5144.
  181. A. T. Phan, Y. S. Modi, D. J. Patel, *J. Am. Chem. Soc.* **2004**, *126*, 8710-8716.
  182. A. T. Phan, V. Kuryavyy, H. Y. Gaw, D. J. Patel, *Nat. Chem. Biol.* **2005**, *1*, 167-173.
  183. S. Amrane, M. Adrian, B. Heddi, A. Serero, A. Nicolas, J.-L. Mergny, A. T. Phan, *J. Am. Chem. Soc.* **2012**, *134*, 5807-5816.
  184. N. Q. Do, K. W. Lim, M. H. Teo, B. Heddi, A. T. Phan, *Nucleic Acids Res.* **2011**, *39*, 9448-9457.
  185. A. T. Phan, V. Kuryavyy, J.-B. Ma, A. Faure, M.-L. Andréola, D. J. Patel, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 634-639.
  186. N. Q. Do, A. T. Phan, *Chem. --Eur. J.* **2012**, *18*, 14752-14759.
  187. D. J. Patel, A. E. Tonelli, *Biopolymers* **1974**, *13*, 1943-1964.
  188. Y. Wang, D. J. Patel, *Biochemistry* **1992**, *31*, 8112-8119.
  189. P. Sket, J. Plavec, *J. Am. Chem. Soc.* **2010**, *132*, 12724-12732.
  190. R. Ida, G. Wu, *J. Am. Chem. Soc.* **2008**, *130*, 3590-3602.
  191. A. T. Phan, D. J. Patel, *J. Am. Chem. Soc.* **2002**, *124*, 1160-1161.
  192. B. Islam, M. Sgobba, C. Laughton, M. Orozco, J. Šponer, S. Neidle, S. Haider, *Nucleic Acids Res.* **2013**, *41*, 2723-2735.
  193. R. Stefl, T. E. Cheatham, N. Spacková, E. Fadrná, I. Berger, J. Koca, J. Šponer, *Biophys. J.* **2003**, *85*, 1787-1804.
  194. E. Fadrná, N. Spacková, R. Stefl, J. Koca, T. E. Cheatham, J. Šponer, *Biophys. J.* **2004**, *87*, 227-242.
  195. A. Pérez, I. Marchán, D. Svozil, J. Šponer, T. E. Cheatham, C. A. Laughton, M. Orozco, *Biophys. J.* **2007**, *92*, 3817-3829.
  196. J. Šponer, N. Špačková, *Methods* **2007**, *43*, 278-290.
  197. F. Fogolari, H. Haridas, A. Corazza, P. Viglino, D. Corà, M. Caselle, G. Esposito, L. E. Xodo, *BMC Struct. Biol.* **2009**, *9*, 64.



198. P. Stadlbauer, L. Trantirek, T. E. Cheatham, 3rd, J. Koca, J. Šponer, *Biochimie* **2014**, *105*, 22-35.
199. P. Stadlbauer, M. Krepl, T. E. Cheatham, 3rd, J. Koca, J. Šponer, *Nucleic Acids Res.* **2013**, *41*, 7128-7143.
200. J. Šponer, X. Cang, T. E. Cheatham, *Methods* **2012**, *57*, 25-39.
201. M. Zgarbova, F. J. Luque, J. Šponer, T. E. Cheatham, 3rd, M. Otyepka, P. Jurecka, *J. Chem. Theory Comput.* **2013**, *9*, 2339-2354.
202. K. Gkionis, H. Kruse, J. A. Platts, A. Mládek, J. Koča, J. Šponer, *J. Chem. Theory Comput.* **2014**, *10*, 1326-1340.
203. H. T. Le, W. L. Dean, R. Buscaglia, J. B. Chaires, J. O. Trent, *J. Phys. Chem. B* **2014**, *118*, 5390-5405.
204. M. Vorlíčková, I. Kejnovská, J. Sagi, D. Renčiuk, K. Bednářová, J. Motlová, J. Kypr, *Methods* **2012**, *57*, 64-75.
205. A. I. Karsisiotis, N. M. Hessari, E. Novellino, G. P. Spada, A. Randazzo, M. Webba da Silva, *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 10645-10648.
206. J. Kypr, I. Kejnovská, D. Renčiuk, M. Vorlíčková, *Nucleic Acids Res.* **2009**, *37*, 1713-1725.
207. S. Paramasivan, I. Rujan, P. H. Bolton, *Methods* **2007**, *43*, 324-331.
208. E. Largy, A. Marchand, V. Gabelica, J.-L. Mergny, *unpublished*.
209. J. L. Mergny, L. Lacroix, *Oligonucleotides* **2003**, *13*, 515-537.
210. J. L. Mergny, L. Lacroix, *Curr. Protoc. Nucleic Acid Chem.* **2009**, Chapter 17, Unit 17.11.
211. A. De Cian, L. Guittat, M. Kaiser, B. Saccà, S. Amrane, A. Bourdoncle, P. Alberti, M. P. Teulade-Fichou, L. Lacroix, J. L. Mergny, *Methods* **2007**, *42*, 183-195.
212. J. L. Mergny, J. C. Maurizot, *Chembiochem* **2001**, *2*, 124-132.
213. M. C. Miller, J. O. Trent, *Curr. Protoc. Nucleic Acid Chem.* **2011**, Chapter 17, Unit 17.13.
214. H. T. Le, M. C. Miller, R. Buscaglia, W. L. Dean, P. A. Holt, J. B. Chaires, J. O. Trent, *Org. Biomol. Chem.* **2012**, *10*, 9393-9404.
215. C. C. Hardin, E. Henderson, T. Watson, J. K. Prosser, *Biochemistry* **1991**, *30*, 4460-4472.
216. T. Laue, *Curr. Opin. Struct. Biol.* **2001**, *11*, 579-583.
217. L. Petraccone, C. Spink, J. O. Trent, N. C. Garbett, C. S. Mekmaysy, C. Giancola, J. B. Chaires, *J. Am. Chem. Soc.* **2011**, *133*, 20951-20961.
218. M. Trajkovski, M. W. da Silva, J. Plavec, *J. Am. Chem. Soc.* **2012**, *134*, 4132-4141.
219. H. T. Le, R. Buscaglia, W. L. Dean, J. B. Chaires, J. O. Trent, *Top. Curr. Chem.* **2013**, *330*, 179-210.
220. U. Dornberger, J. Behlke, E. Birch-Hirschfeld, H. Fritzsche, *Nucleic Acids Res.* **1997**, *25*, 822-829.
221. C. S. Mekmaysy, L. Petraccone, N. C. Garbett, P. A. Ragazzon, R. Gray, J. O. Trent, J. B. Chaires, *J. Am. Chem. Soc.* **2008**, *130*, 6710-6711.
222. C. Rehm, I. T. Holder, A. Groß, F. Wojciechowski, M. Urban, M. Sinn, M. Drescher, J.

- S. Hartig, *Chem. Sci.* **2014**, *5*, 2809-2809.
223. M. C. Miller, H. T. Le, W. L. Dean, P. A. Holt, J. B. Chaires, J. O. Trent, *Org. Biomol. Chem.* **2011**, *9*, 7633-7637.
  224. G. W. Collie, G. N. Parkinson, S. Neidle, F. Rosu, E. De Pauw, V. Gabelica, *J. Am. Chem. Soc.* **2010**, *132*, 9328-9334.
  225. F. Rosu, V. Gabelica, H. Poncelet, E. De Pauw, *Nucleic Acids Res.* **2010**, *38*, 5217-5225.
  226. V. Gabelica, E. Schulz, M. Karas, *J. Mass Spectrom.* **2004**, *39*, 579-593.
  227. V. Gabelica, E. De Pauw, *Mass. Spectrom. Rev.* **2005**, *24*, 566-587.
  228. A. Marchand, V. Gabelica, *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 1146-1154.
  229. A. Marchand, A. Granzhan, K. Iida, Y. Tsushima, Y. Ma, K. Nagasawa, M. P. Teulade-Fichou, V. Gabelica, *J. Am. Chem. Soc.* **2015**, *137*, 750-756.
  230. F. Lanucara, S. W. Holman, C. J. Gray, C. E. Eyers, *Nature Chem.* **2014**, *6*, 281-294.
  231. R. Ferreira, A. Marchand, V. Gabelica, *Methods* **2012**, *57*, 56-63.
  232. *Nucleic Acids in the Gas Phase*, Ed. V. Gabelica, Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.
  233. K. Okamoto, Y. Sannohe, T. Mashimo, H. Sugiyama, M. Terazima, *Bioorg. Med. Chem.* **2008**, *16*, 6873-6879.
  234. R. A. Darby, M. Sollogoub, C. McKeen, L. Brown, A. Risitano, N. Brown, C. Barton, T. Brown, K. R. Fox, *Nucleic Acids Res.* **2002**, *30*, e39-e39.
  235. A. Risitano, K. R. Fox, *Nucleic Acids Res.* **2004**, *32*, 2598-2606.
  236. E. E. Merkina, K. R. Fox, *Biophys. J.* **2005**, *89*, 365-373.
  237. S. Redon, S. Bombard, M. A. Elizondo-Riojas, J. C. Chottard, *Nucleic Acids Res.* **2003**, *31*, 1605-1613.
  238. H. Bertrand, S. Bombard, D. Monchaud, E. Talbot, A. Guédin, J.-L. Mergny, R. Grünert, P. J. Bednarski, M.-P. Teulade-Fichou, *Org. Biomol. Chem.* **2009**, *7*, 2864-2871.
  239. Y. He, R. D. Neumann, I. G. Panyutin, *Nucleic Acids Res.* **2004**, *32*, 5359-5367.
  240. T. I. Gaynutdinov, R. D. Neumann, I. G. Panyutin, *Nucleic Acids Res.* **2008**, *36*, 4079-4087.
  241. J. E. Redman, *Methods* **2007**, *43*, 302-312.
  242. R. Hänsel, F. Loehr, L. Trantirek, V. Doetsch, *J. Am. Chem. Soc.* **2013**, *135*, 2816-2824.
  243. J. Y. Lee, J. Yoon, H. W. Kihm, D. S. Kim, *Biochemistry* **2008**, *47*, 3389-3396.
  244. T. Miura, J. M. Benevides, G. J. Thomas, *J. Mol. Biol.* **1995**, *248*, 233-238.
  245. R. M. Abu-Ghazalah, S. Rutledge, L. W. Y. Lau, D. N. Dubins, R. B. MacGregor, A. S. Helmy, *Biochemistry* **2012**, *51*, 7357-7366.
  246. S. Ceru, P. Sket, I. Prislan, J. Lah, J. Plavec, *Angew. Chem. Int. Ed. Engl.* **2014**, 1002-1002.
  247. B. Pagano, C. A. Mattia, C. Giancola, *Int. J. Mol. Sci.* **2009**, *10*, 2935-2957.
  248. B. Pagano, A. Randazzo, I. Fotticchia, E. Novellino, L. Petraccone, C. Giancola, *Methods* **2013**, *64*, 43-51.
  249. N. M. Brown, P. A. Rachwal, T. Brown, K. R. Fox, *Org. Biomol. Chem.* **2005**, *3*, 4153-

- 4157.
250. P. A. Rachwal, I. S. Findlow, J. M. Werner, T. Brown, K. R. Fox, *Nucleic Acids Res.* **2007**, *35*, 4214-4222.
  251. S. Neidle, *Methods* **2007**, *43*, 245.
  252. D. Sun, L. H. Hurley, *Methods Mol. Biol.* **2010**, *608*, 65-79.
  253. C. Detellier, P. Laszlo, *J. Am. Chem. Soc.* **1980**, *102*, 1135-1141.
  254. M. Borzo, C. Detellier, P. Laszlo, A. Paris, *J. Am. Chem. Soc.* **1980**, *102*, 1124-1134.
  255. A. Wong, R. Ida, G. Wu, *Biochem. Biophys. Res. Commun.* **2005**, *337*, 363-366.
  256. Q. Xu, H. Deng, W. H. Braunlin, *Biochemistry* **1993**, *32*, 13130-13137.
  257. H. Deng, W. H. Braunlin, *J. Mol. Biol.* **1996**, *255*, 476-483.
  258. G. Wu, A. Wong, Z. Gan, J. T. Davis, *J. Am. Chem. Soc.* **2003**, *125*, 7182-7183.
  259. M. L. Gill, S. A. Strobel, J. P. Loria, *J. Am. Chem. Soc.* **2005**, *127*, 16723-16732.
  260. P. Sket, M. Črnugelj, J. Plavec, *Nucleic Acids Res.* **2005**, *33*, 3691-3697.
  261. K. Snoussi, B. Halle, *Biochemistry* **2008**, *47*, 12219-12229.
  262. F. Rosu, V. Gabelica, C. Houssier, P. Colson, E. De Pauw, *Rapid Commun. Mass Spectrom.* **2002**, *16*, 1729-1736.
  263. J. Gros, F. Rosu, S. Amrane, A. De Cian, V. Gabelica, L. Lacroix, J.-L. Mergny, *Nucleic Acids Res.* **2007**, *35*, 3064-3075.
  264. L. Joly, F. Rosu, V. Gabelica, *Chem. Commun.* **2012**, *48*, 8386-8388.
  265. A. Marchand, R. Ferreira, H. Tateishi-Karimata, D. Miyoshi, N. Sugimoto, V. Gabelica, *J. Phys. Chem. B* **2013**, *117*, 12391-12401.
  266. F. Balthasart, J. Plavec, V. Gabelica, *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 1-8.
  267. R. D. Gray, J. B. Chaires, *Methods* **2012**, *57*, 47-55.
  268. I. Bang, *Biochem. Zeitschrift* **1910**, *26*, 293-311.
  269. M. Gellert, M. N. Lipsett, D. R. Davies, *Proc. Natl. Acad. Sci. U. S. A.* **1962**, *48*, 2013-2018.
  270. P. Tougard, J. F. Chantot, W. Guschlbauer, *Biochim. Biophys. Acta* **1973**, *308*, 9-16.
  271. H. Q. Yu, D. Miyoshi, N. Sugimoto, *J. Am. Chem. Soc.* **2006**, *128*, 15461-15468.
  272. J. Kim, C. Cheong, P. B. Moore, *Nature* **1991**, *351*, 331-332.
  273. Y. Krishnan-Ghosh, E. Stephens, S. Balasubramanian, *J. Am. Chem. Soc.* **2004**, *126*, 5944-5945.
  274. E. Gavathiotis, M. S. Searle, *Org. Biomol. Chem.* **2003**, *1*, 1650-1656.
  275. V. L. Makarov, Y. Hirose, J. P. Langmore, *Cell* **1997**, *88*, 657-666.
  276. R. K. Moyzis, J. M. Buckingham, L. S. Cram, M. Dani, L. L. Deaven, M. D. Jones, J. Meyne, R. L. Ratliff, J. R. Wu, *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 6622-6626.
  277. C. C. Hardin, M. J. Corregan, D. V. Lieberman, B. A. Brown, 2nd, *Biochemistry* **1997**, *36*, 15428-15450.
  278. J. R. Wyatt, P. W. Davis, S. M. Freier, *Biochemistry* **1996**, *35*, 8002-8008.
  279. C. Cheong, P. B. Moore, *Biochemistry* **1992**, *31*, 8406-8414.
  280. M. K. Raghuraman, T. R. Cech, *Nucleic Acids Res.* **1990**, *18*, 4543-4552.
  281. J. J. Green, L. Ying, D. Klennerman, S. Balasubramanian, *J. Am. Chem. Soc.* **2003**, *125*, 3763-3767.

282. T. Mashimo, Y. Sannohe, H. Yagi, H. Sugiyama, *Nucleic Acids Symp. Ser.* **2008**, 409-410.
283. T. Mashimo, H. Sugiyama, *Nucleic Acids Symp. Ser.* **2007**, 239-240.
284. T. Mashimo, H. Yagi, Y. Sannohe, A. Rajendran, H. Sugiyama, *J. Am. Chem. Soc.* **2010**, *132*, 14910-14918.
285. R. D. Gray, J. O. Trent, J. B. Chaires, *J. Mol. Biol.* **2014**, *426*, 1629-1650.
286. R. Buscaglia, R. D. Gray, J. B. Chaires, *Biopolymers* **2013**, *99*, 1006-1018.
287. R. D. Gray, R. Buscaglia, J. B. Chaires, *J. Am. Chem. Soc.* **2012**, *134*, 16834-16844.
288. R. D. Gray, J. Li, J. B. Chaires, *J. Phys. Chem. B* **2009**, *113*, 2676-2683.
289. M. Bončina, J. Lah, I. Prislan, G. Vesnaver, *J. Am. Chem. Soc.* **2012**, *134*, 9657-9663.
290. N. An, A. M. Fleming, C. J. Burrows, *J. Am. Chem. Soc.* **2013**, *135*, 8562-8570.
291. N. An, A. M. Fleming, E. G. Middleton, C. J. Burrows, *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 14325-14331.
292. A. Y. Zhang, S. Balasubramanian, *J. Am. Chem. Soc.* **2012**, *134*, 19297-19308.
293. 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.
294. L. Cerofolini, J. Amato, A. Giachetti, V. Limongelli, E. Novellino, M. Parrinello, M. Fragai, A. Randazzo, C. Luchinat, *Nucleic Acids Res.* **2014**, *42*, 13393-13404.
295. J. Töhl, W. Eimer, *Biophys. Chem.* **1997**, *67*, 177-186.
296. N. V. Hud, J. Plavec, in *Quadruplex Nucleic Acids*, Eds. S. Neidle, S. Balasubramanian, Royal Society of Chemistry, Cambridge, UK, 2006, pp. 301.
297. L. Ma, M. Iezzi, M. S. Kaucher, Y.-F. Lam, J. T. Davis, *J. Am. Chem. Soc.* **2006**, *128*, 15269-15277.
298. A. E. Engelhart, J. Plavec, Ö. Persil, N. V. Hud, in *Nucleic Acid-Metal Ion Interactions*, Eds. N. V. Hud, S. Neidle, Royal Society of Chemistry, Cambridge, UK, 2009, pp. 433.
299. P. Sket, A. Virgilio, V. Esposito, A. Galeone, J. Plavec, *Nucleic Acids Res.* **2012**, *40*, 11047-11057.
300. R. V. Reshetnikov, J. Šponer, O. I. Rassokhina, A. M. Kopylov, P. O. Tsvetkov, A. A. Makarov, A. V. Golovin, *Nucleic Acids Res.* **2011**, *39*, 9789-9802.
301. D. Sen, W. Gilbert, *Curr. Opin. Struct. Biol.* **1991**, *1*, 435-438.
302. R. M. Brosh, *Aging* **2011**, *3*, 332-335.
303. M. Dong, T. Mürdter, U. Klotz, *Eur. J. Clin. Pharmacol.* **2010**, *66*, 1-3.
304. A. De Cian, L. Lacroix, C. Douarre, N. Temime-Smaali, C. Trentesaux, J.-F. Riou, J.-L. Mergny, *Biochimie* **2008**, *90*, 131-155.
305. V. Gabelica, E. S. Baker, E. D. Pauw, M.-P. Teulade-Fichou, E. De Pauw, M. T. Bowers, *J. Am. Chem. Soc.* **2007**, *129*, 895-904.
306. D. Renčiuk, I. Kejnovská, P. Skoláková, K. Bednářová, J. Motlová, M. Vorlíčková, *Nucleic Acids Res.* **2009**, *37*, 6625-6634.
307. V. Viglasky, L. Bauer, K. Tluczkova, P. Javorsky, *J. Nucleic Acids* **2010**, *2010*, Article ID 820356.
308. I. N. Rujan, J. C. Meleney, P. H. Bolton, *Nucleic Acids Res.* **2005**, *33*, 2022-2031.

309. M. C. Miller, R. Buscaglia, J. B. Chaires, A. N. Lane, J. O. Trent, *J. Am. Chem. Soc.* **2010**, *132*, 17105-17107.
310. R. D. Gray, L. Petraccone, J. O. Trent, J. B. Chaires, *Biochemistry* **2010**, *49*, 179-194.
311. Y. Xu, Y. Noguchi, H. Sugiyama, *Bioorg. Med. Chem.* **2006**, *14*, 5584-5591.
312. 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*, 7934-7946.
313. J. Li, J. J. Correia, L. Wang, J. O. Trent, J. B. Chaires, *Nucleic Acids Res.* **2005**, *33*, 4649-4659.
314. R. Hänsel, F. Löhr, S. Foldynová-Trantírková, E. Bamberg, L. Trantírek, V. Dötsch, *Nucleic Acids Res.* **2011**, *39*, 5768-5775.
315. D. Sen, W. Gilbert, *Nature* **1990**, *344*, 410-414.
316. M. A. Keniry, G. D. Strahan, E. A. Owen, R. H. Shafer, *Eur. J. Biochem.* **1995**, *233*, 631-643.
317. G. D. Strahan, M. a. Keniry, R. H. Shafer, *Biophys. J.* **1998**, *75*, 968-981.
318. M. Črnugelj, N. V. Hud, J. Plavec, *J. Mol. Biol.* **2002**, *320*, 911-924.
319. M. Črnugelj, P. Sket, J. Plavec, *J. Am. Chem. Soc.* **2003**, *125*, 7866-7871.
320. F. W. Smith, P. Schultze, J. Feigon, *Structure* **1995**, *3*, 997-1008.
321. C. C. Hardin, T. Watson, M. Corregan, C. Bailey, *Biochemistry* **1992**, *31*, 833-841.
322. W. Li, X. M. Hou, P. Y. Wang, X. G. Xi, M. Li, *J. Am. Chem. Soc.* **2013**, *135*, 6423-6426.
323. D. Sen, W. Gilbert, *Nature* **1988**, *334*, 364-366.
324. A. Tanaka, J. Choi, T. Majima, *RSC Adv.* **2014**, *4*, 59071-59077.
325. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605-1612.