# PHYSICO-CHEMICAL CHARACTERIZATION OF DNA G-QUADRUPLEXES AND THEIR INTERACTION WITH PROTEINS AND POTENTIAL ANTICANCER AGENTS

Bruno Pagano

Dottorato di Ricerca in Scienze Farmaceutiche

Università degli Studi di Salerno

UNIVERSITÀ DEGLI STUDI DI SALERNO
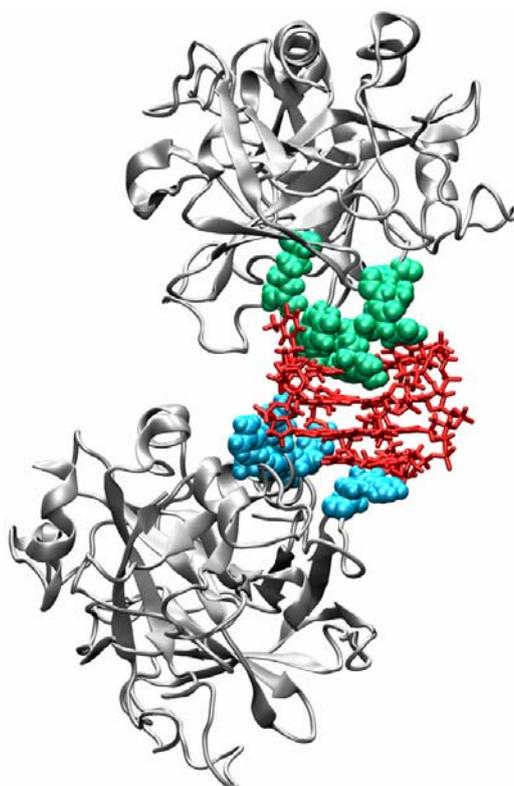
Dottorato di Ricerca in Scienze Farmaceutiche

VI-CICLO NUOVA SERIE

2004-2007

CARATTERIZZAZIONE CHIMICO-FISICA DI QUADRUPLE ELICHE DI DNA E LORO INTERAZIONE CON PROTEINE E POTENZIALI AGENTI ANTITUMORALI

Dottorando
Bruno Pagano

Coordinatore
Prof. Nunziatina De Tommasi

Tutor
Prof. Carlo A. Mattia

UNIVERSITÀ DEGLI STUDI DI SALERNO

Unione Europea

Ministero dell'Istruzione,
dell'Università e della Ricerca

RIASSUNTO

Negli ultimi anni è stato dimostrato che sequenze di DNA (o RNA) ricche in guanine sono in grado di formare una nuova classe di strutture: le quadruple eliche. Importante e' la presenza di quadruple eliche in regioni promotrici e regolatrici di molti geni e nella parte terminale dei cromosomi, i telomeri.

Recentemente, lo studio di queste molecole ha avuto un notevole sviluppo grazie alle diverse applicazioni che esse potrebbero avere in campo medico e farmaceutico. In particolare, le quadruple eliche hanno una potenziale applicazione nella terapia anticancro quali inibitori della telomerasi, un enzima la cui iperattività è sicuramente collegata allo sviluppo del cancro. Un promettente approccio per inibire l'attività della telomerasi riguarda l'utilizzo di agenti che possano legare e stabilizzare le quadruple eliche presenti ai telomeri bloccando, in tal modo, l'attività catalitica dalla telomerasi ed agendo da antitumorali.

Inoltre, è stato scoperto che oligonucleotidi con struttura a quadrupla elica possono agire da aptameri, cioè hanno la capacità di legare specificamente delle proteine bersaglio, inibendole. L'aptamero denominato TBA *(Thrombin Binding Aptamer)* è un oligomero a DNA con struttura a quadrupla elica, scoperto *in vitro*, che è in grado di legare la trombina inibendone la funzione. Il TBA per le sue proprietà anticoagulanti è stato oggetto di numerosi studi sia di tipo strutturale sia di natura farmacologica. L'attività anticoagulante del TBA *in vitro* è tale da giustificare un suo eventuale impiego in terapia, ma purtroppo tale attività risulta piuttosto bassa *in vivo* poichè il DNA è rapidamente degradato dalle nucleasi.

Uno degli obiettivi di questa tesi di dottorato è stato quello di affrontare uno studio chimico-fisico di composti che sono in grado di legarsi in maniera specifica alle quadruple eliche e dei fattori che regolano gli equilibri in gioco; tali composti sono, infatti, in grado di incrementare la stabilità di questi sistemi e di conseguenza ne possono aumentare l'effetto terapeutico. Per questo motivo è stata studiata la termodinamica dell'interazione di alcune molecole di interesse farmacologico, quali la distamicina, due suoi derivati e la porfirina cationica, con diverse quadruple eliche. Le quadruple eliche prese in esame sono state la $d[AG_3(TTAG_3)_3]$ e la $[d(TAGGGTTAGGG)]_2$ che rappresentano due quadruple eliche formate da un diverso troncamento dal DNA telomerico umano e la $d[(TGGGGT)]_4$ formata da una sequenza troncata del DNA telomerico di *Oxytricha e Tetrahymena*.

Un altro obiettivo è stato quello di caratterizzare la stabilità termodinamica del TBA e di un aptamero modificato (mTBA) ed il processo di *binding* di questi aptameri con la trombina. L'mTBA presenta una modifica chimica in grado di rendere resistente l'aptamero all'azione delle

nucleasi che potrebbe rendere concreto un suo eventuale utilizzo come principio attivo di un farmaco anticoagulante.

Inoltre, è stato anche affrontato uno studio computazionale di due quadruple eliche bimolecolari formate da sequenze analoghe di RNA e DNA allo scopo di chiarire i motivi per cui si formano due diverse strutture a quadrupla elica ed i fattori che le stabilizzano.

Lo studio è stata condotto principalmente attraverso tecniche calorimetriche quali la calorimetria isoterma (ITC), la microcalorimetria differenziale a scansione (DSC), il dicroismo circolare (CD), e metodi computazionali quali meccanica e dinamica molecolare e *docking*.

I risultati ottenuti per i sistemi quadrupla elica-ligando dimostrano che la distamicina ed un suo derivato contenente un anello metil pirrolico in più ed un gruppo ammidico terminale, legano le quadruple eliche in esame, al contrario un secondo derivato della distamicina contenente due anelli metil pirrolici in più ed un gruppo ammidico terminale non ha mostrato avere interazioni specifiche.

I parametri termodinamici ottenuti indicano che le interazioni quadrupla elica-ligando sono fortemente influenzate dallo ione presente in soluzione e che anche la stechiometria di legame è dipendente dal tipo di soluzione. Il calcolo dei parametri termodinamici ha mostrato che, sia in sodio che in potassio, il legame della distamicina e del composto I alle quadruple eliche in esame è entropicamente guidato. Questo dato in particolare suggerisce che il *binding* di queste molecole potrebbe avvenire nei solchi delle quadruple eliche. Inoltre, un risultato particolarmente interessante è che il composto I ha mostrato, in entrambe le soluzioni, una maggiore affinità della distamicina per le quadruple eliche.

I risultati ottenuti per l'interazione porfirina-[d(TAGGGTTAGGG)]$_2$ sono in generale accordo con la struttura cristallografica, recentemente riportata, del complesso ed indicano un *binding* entalpicamente guidato con una stechiometria 2:1.

Lo studio della stabilità termodinamica del TBA e del TBA modificato ha rivelato che l'introduzione della modifica aumenta sia la stabilità termica che termodinamica dell'aptamero. Inoltre, i valori delle variazioni di entalpia ed entropia per la dissociazione del TBA modificato risultano più elevati rispetto al TBA, suggerendo una struttura più rigida e la presenza di ulteriori interazioni intramolecolari nell'aptamero modificato. Questi risultati sono stati confermati ed interpretati sulla base dei risultati delle dinamiche molecolari dei due aptameri. I parametri termodinamici del processo di *binding* del TBA e del TBA modificato con la trombina mostrano che l'interazione è esotermica e che la stechiometria del complesso che si forma è 1:2 (aptamero:proteina). Il TBA modificato risulta però avere una più alta affinità per la trombina e la sua interazione è associata ad una variazione di entalpia maggiormente favorevole. I risultati delle

dinamiche molecolari dei complessi suggeriscono che la maggior affinità dell'aptamero modificato sia dovuta essenzialmente ad una migliore interazione con l'esosito II della trombina.

Lo studio di dinamica molecolare delle quadruple eliche di RNA e DNA ha evidenziato che la maggiore differenza tra le due molecole deriva dalla presenza del gruppo 2'-OH del ribosio, che contribuisce considerevolmente al numero di legami idrogeno formando più interazioni intramolecolari nella molecola di RNA. I legami che hanno una maggiore persistenza sono quelli formati con l'ossigeno del gruppo fosfato. Questo tipo di legame idrogeno conferisce rigidità alla struttura ed è in parte responsabile della stabilità delle esadi presenti nella quadrupla elica di RNA. Inoltre, durante le simulazioni, gli ioni Na^+, inizialmente non direttamente coordinati alle quadruple eliche, vanno a coordinare in modo stabile le macromolecole indicando un possibile meccanismo di coordinazione non ancora osservato sperimentalmente.

Parte della ricerca è stata svolta in collaborazione con la Prof. Concetta Giancola del Dipartimento di Chimica dell'Università Federico II di Napoli. Inoltre, una parte dello studio è stato realizzato a Londra presso i laboratori del Prof. Stephen Neidle della School of Pharmacy dell'University of London, e presso il gruppo della Dott. Franca Fraternali, della Randall Division del King's College London.

CHAPTER 1: INTRODUCTION

1.1 General overview of nucleic acids

Nucleic acids are the biological macromolecules that store and transmit genetic information in living organisms. This information ensures the normal development and functioning of an organism. Nucleic acids comprise of two main classes: ribonucleic acids (RNA) and deoxyribonucleic acids (DNA). The structure of DNA plays an essential role in DNA's ability to store and transmit genetic information. DNA is a polymer of simple units, and the genetic information is encoded in the sequence of these monomeric units, the nucleotides. Nucleotides combine together to form DNA strands, and complementary strands of DNA bind together to form a double stranded helix. The integrity of the genetic information is maintained within this DNA structure. This information is converted via several processes (transcription and translation) into various structural, regulatory and functional components *in vivo* [1]. Consequently, understanding the factors that influence the structure of DNA is significant for understanding the processes that compromise this information such as through DNA damage.

A nucleotide is a chemical compound that consists of three components: a heterocyclic base, a sugar and one o more phosphate groups. The bases of DNA are divided into two groups: purines [adenine (A) and guanine (G)] and pyrimidines [thymine (T) and cytosine (C)]. In RNA, thymine is replaced by uracil (U). Purines consist of a six-membered and a five-membered nitrogen-containing ring, fused together; pyridmidines have only a six-membered nitrogen-containing ring (**Fig. 1.1**) [1].
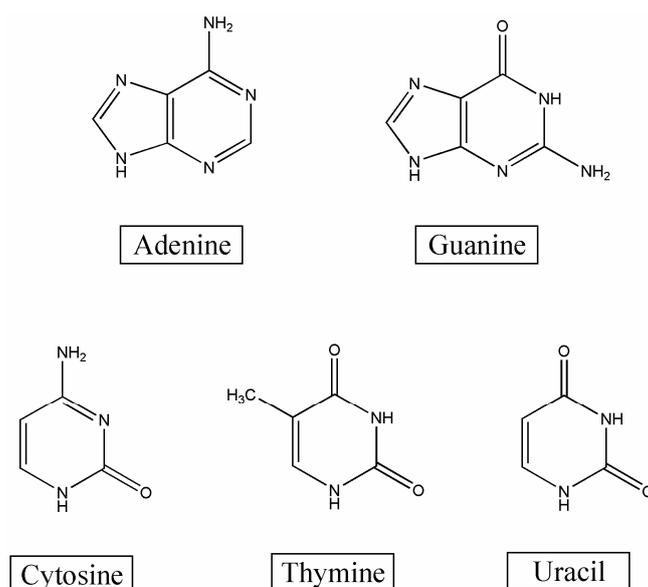

Fig. 1.1 – *Structure of purines and pyrimidines.*

The sugar component of nucleic acids is a pentose, DNA contains the 2'-deoxyribose, while RNA contains ribose. The molecule comprising the pentose sugar joined to a base is called nucleoside. In a nucleoside, sugar and base are covalently bound by a N-glycosidic bond between the C1' of the sugar and the N9 or N1 of purines or pyrimidines, respectively. Rotations about this glycosidic linkage add structural diversity to the DNA molecule. The two main conformations that are formed about the glycosidic bond are either *syn* or *anti* (**Fig. 1.2**). The *syn* conformation is formed when the C1'-O4' bond is cis to the N9-C4 purine bond, and cis to the N1-C2 of pyrimidine. This occurs when the bulk of the purine base is facing towards the sugar or the C2 carbonyl is on top of the sugar ring. The *anti* conformation is formed when the bulk of the base is rotated away from the sugar. This occurs when the C1'-O4' bond of the sugar is trans to the N9-C4 bond of the purine base. In pyrimidines, the *anti* conformation is formed when the C2 carbonyl faces away from the sugar. This occurs when the C1'-O4' bond is trans to the N1-C2 pyrimidine bond.
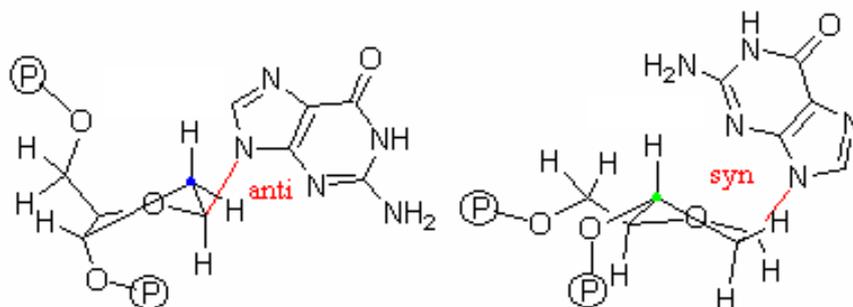
Fig. 1.2 – *Rotation about the glycosidic bond.*

A nucleoside is converted into a nucleotide by attachment of a phosphate group to the C5' carbon of the sugar. The nucleotides are connected by phosphodiester bonds to form single-stranded DNA.

The phosphodiester bond is formed by the 5' phosphate of one nucleotide attaching to the 3' hydroxyl of the next nucleotide. This attachment lends directionality to DNA. The DNA sequence is usually read from the 5' to 3' direction. The 5'-end commonly has a phosphate group attached to it while the 3'-end contains just the sugar hydroxyl.

The bases of DNA (or RNA) can form non-covalent bond couplings called base pairs. The most common base pairing pattern found in DNA is the Watson-Crick base pairing. In the canonical Watson-Crick base pairing, adenine forms a base pair with thymine, as does guanine with cytosine in DNA [2]. The base pairs are held together by hydrogen bonds between covalently bound hydrogen atoms and hydrogen bond acceptor atoms such as carbonyl oxygen atoms or nitrogen

atoms. Adenine and thymine (A:T) base pairs are held together by two hydrogen bonds, whereas guanine and cytosine (G:C) base pairs are held by three hydrogen bonds. Although Watson-Crick base pairing is more commonly found in DNA, Hoogsteen base pairing has been observed in several DNA structures such as parallel, triplex and quadruplex DNA structures. Hoogsteen base pairing alters the normal base pairing pattern by base pairing through the N7 position.

Usually, DNA assumes a double stranded helix structure (DNA duplex). The most common DNA duplex conformation is the B-form, which is a right-handed helix [2]. The two complementary strands are anti-parallel, one strand is oriented in the 5'-3' direction while the other strand is oriented in the 3'-5' direction. The bases are within the helix, forming base-pairs that are perpendicular to the axis of the helix. B-DNA has two distinct grooves within its edges due to the orientation of the base pairs: a major groove and a minor groove. The double helix is stabilized and held together by several different forces. Firstly, hydrogen bonding between the bases on the complementary strands stabilizes the DNA duplex. Secondly, hydrophobic base-stacking interactions between the aromatic rings of adjacent base pairs also contribute to stability. Thirdly, water molecules cooperatively bind along the grooves of the DNA adding further stability to the duplex structure. Finally, metal cations surround the negatively charged phosphate groups in the sugar-phosphate backbone adding more stability to the DNA duplex. Over and above B-form, DNA duplex can also assumes, under different conditions, a number of other different conformations such as the A-form and Z-form [1].

1.2 Other DNA structures

1.2.1 Triplex

The existence of the DNA triple helix structure is known since 1957 [3], four years after the discovery of the DNA double helix structure [2], but only recently the triple-helical forms of DNA have received the due attention. The interest in triple helices of DNA is in large part due to the potential biomedical applications for specific controlling gene expression at both transcriptional and replicational levels [4].

Canonical triple helix formation relies upon hydrogen bonding interaction between a homopyrimidine oligonucleotide and homopurine-homopyrimidine duplex already engaged in Watson-Crick hydrogen bonds. The third oligonucleotide strand (often called TFO, acronym of Triplex Forming Oligonucleotide) occupies the major groove of the double helix forming Hoogsteen hydrogen bonds [5] with the purines of the Watson-Crick type base pairs. This recognition process is extremely simple, specific and efficient. Specificity is derived from thymine (T) recognition of adenine-thymine base pair (A·T) and protonated cytosine (C^+) recognition of guanine-cytosine base pair (G·C). The efficiency of the single strand binding depends on several factors, such as pH, cation concentration, composition and length of the third strand. NMR [6], infrared spectroscopy [7] and fiber diffraction studies [8] convincingly demonstrated that the original double helix within the triplex adopts a B-like rather than A-like structure as was initially concluded by Arnott and Selsing in 1974 [9].

1.2.2 G-Quadruplex

G-quadruplexes are higher-order DNA and RNA structures formed from G-rich sequences that are stabilized by tetrads of hydrogen-bonded guanine bases (G-quartets) (**Fig. 1.3**) [10]. Stacking of G-quartets and coordination of metal cations, preferentially Na^+ and K^+, between two adjacent G-quartet planes yield a stable G-quadruplex structure [11]. Variations in the molecularity, topology, strand orientation, and glycosidic conformation of the G-quadruplex DNA provide a diverse array of structures [12].
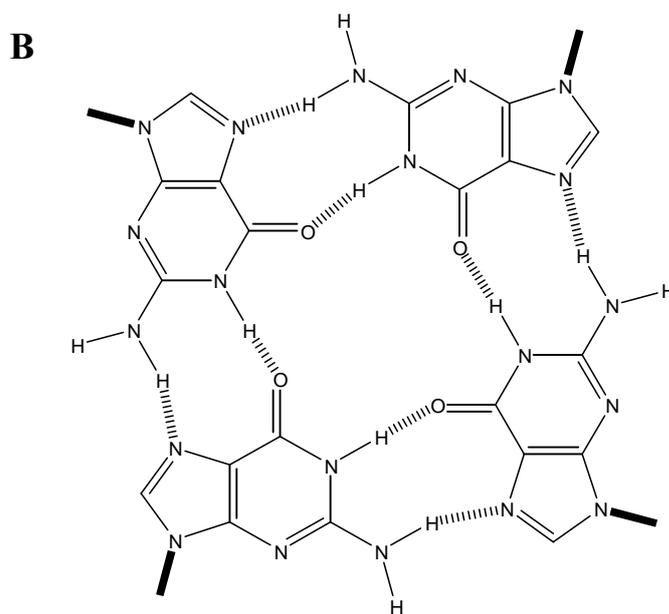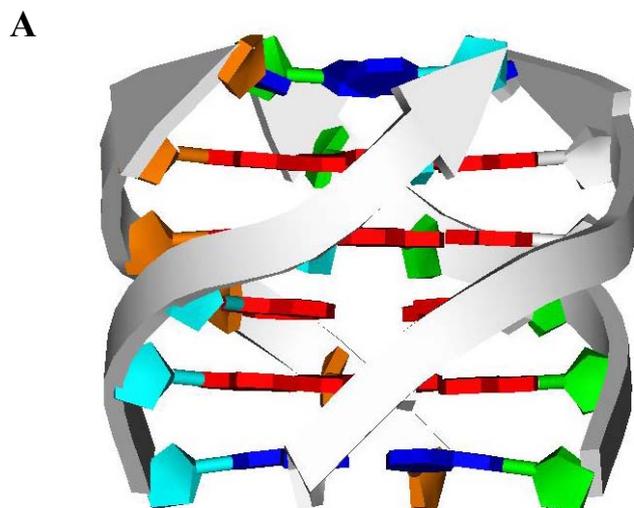
Fig. 1.3 – *(A) Example of G-quadruplex structure. Guanine bases forming the G-tetrads are highlighted in red; (B) top view of a G-tetrad.*

1.3 G-Quadruplex structures

1.3.1 Fundamentals of G-quadruplex structures

G-quadruplexes can be formed from one, two or four separate G-reach strands of DNA (or RNA) and can display a wide variety of topologies, which are in part a consequence of various possible combinations of strand direction, as well as variations in loop size and sequence [13]. They can be defined in general terms as structures formed by a core of at least two stacked G-quartets (or G-tetrads), which are held together by loops that are not involved in the quartets themselves. The G-quartet consists of a planar arrangement of four guanine bases associated through a cyclic array of Hoogsteen-like hydrogen bonds in which each guanine base both accepts and donates two hydrogen bonds (**Fig. 1.3**) [10]. The G-tetrads are not stacked linearly, but adopt a right-handed helix. The combination of the number of stacked G-quartets, the polarity of the strands and the location and length of the loops would be expected to lead to a plurality of G-quadruplex structures, as indeed is found experimentally [12, 14].

G-quadruplexes can be classified on the basis of: 1. the number of strands (one, two or four); 2. the pattern of strand orientation (parallel, antiparallel); 3. the conformation of guanine glycosidic torsion angles (*anti* or *syn*); 4. the orientation of the loops (lateral, diagonal or propeller) (**Fig. 1.4**) (for a recent review see ref. [15]).

Unimolecular quadruplexes are generated by a single oligonucleotide strand whose G-tracts are involved in intramolecular G-tetrad interactions. Potential unimolecular G-quadruplex-forming sequences can be described as follows:

$$G_m X_n G_m X_o G_m X_p G_m$$

where m is the number of G residues in each short G-tract, which are usually directly involved in G-tetrad interactions. X_n, X_o and X_p can be any combination of residues, including G, forming the loops. This notation also implies that the G-tracts can be of unequal length, and if one of the short G tracts is longer than the others, some of the G residues will be located in the loop regions.

Bimolecular and tetramolecular quadruplexes are formed by association of two and four strands, respectively, and, in principle, they can be formed from the association of non-equal sequences. However, almost all bimolecular quadruplexes reported to date are formed by the association of two identical sequences $X_n G_m X_o G_m X_p$, where n and p may or may not be zero. Tetramolecular quadruplexes may be formed by four $X_n G_m X_o$ or $G_m X_n G_m$ strands associating together.

X-ray and NMR structures revealed that all the strands of tetramolecular quadruplexes are in a parallel orientation. All parallel quadruplexes have all guanine glycosidic angles in an *anti* conformation [16, 17]. Quadruplexes are defined antiparallel when at least one of the strands is antiparallel to the others. Antiparallel quadruplexes have both *anti* and *syn* guanine glycosidic

torsion angles and their distribution along the strand depends on distinctive topological arrangements, since different topologies have the four strands in differing positions relative to each other [18-20].
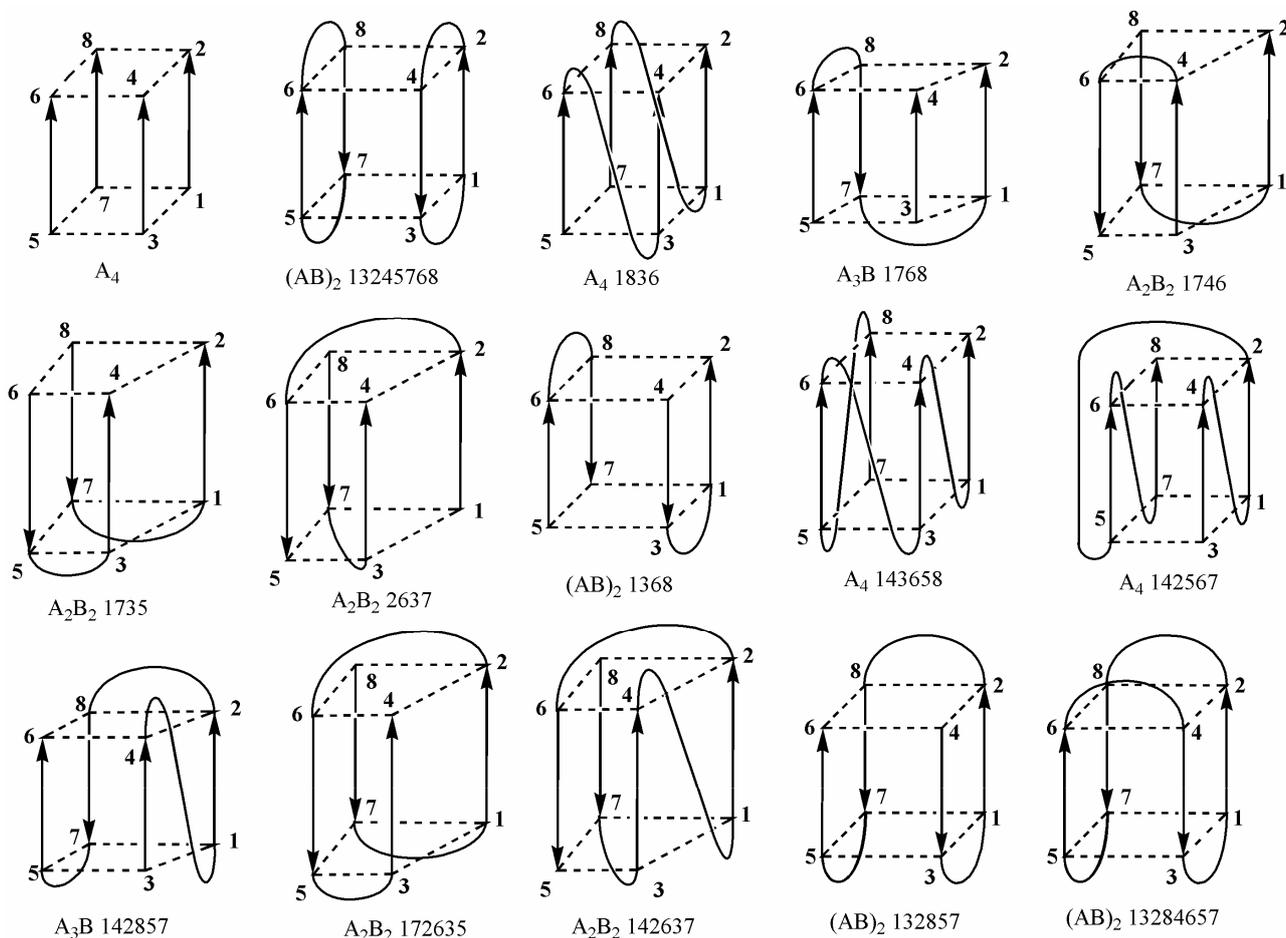
Fig. 1.4 – *Some possible topologies for tetramolecular, bimolecular and unimolecular quadruplexes. The nomenclature indicated is based on a recently proposed topological classification of quadruplexes in which structures are ordered in four main families according to their relative strands orientation (capital letters) and the loops connectivities (arabic numbers)* [21].

Three types of loops have been observed in quadruplex structures: lateral, diagonal or propeller loops. In general, lateral loops join adjacent G-strands, whereas diagonal loops join opposite G-strands; in the third type of loops, propeller loops, the loop joins the bottom G-tetrad with the top G-tetrad [15]. In general, unimolecular quadruplexes present all the three types of loop and a variety of antiparallel strand orientations, although parallel strand orientations were also found [22, 23].

Bimolecular quadruplexes can form different structures depending on even small changes in sequence. For example, the telomeric sequence d(G$_4$T$_4$G$_4$) of the cyliate *Oxytricha nova* forms in solution a symmetric bimolecular structure with parallel/antiparallel strands orientation, diagonal loops formed by opposite strands and alternating *anti*/*syn* guanine glycosidic angles [24-27]. Both sequences d(G$_3$T$_4$G$_4$) and d(G$_4$T$_4$G$_3$) with a missing guanine at the 5'- or 3'-end with respect to the *Oxytricha nova* sequence, form an asymmetric bimolecular quadruplex. However, the first structure shows one lateral and one diagonal loop [28], whereas the second structure shows two diagonal loops [29].

1.3.2 G-Quadruplex groove widths

All quadruplex structures have four grooves, defined as the cavities bounded by the phosphodiester backbones. A consequence of variations in the glycosidic torsion angles is the altering of the spacing between the DNA strands of the G-quadruplex, resulting in varying groove-widths along the side of the quadruplex [30]. In cases where a guanosine in the *syn* conformation donates hydrogen bonds to a neighboring guanosine in the *anti* conformation, the groove formed between the two is extremely narrow, with a phosphate to phosphate distance as small as 7-9 Å [31]. In contrast, when the hydrogen bonding polarity between adjacent *syn*- and *anti*-guanosines is reversed, a very wide groove is formed [31]. An intermediate width groove results when adjacent guanosines adopt the same glycosidic conformation [31]. Indeed, in a parallel four-stranded quadruplex, where all the guanine bases are in the *anti* conformation, the four grooves are approximately the same medium width (**Fig. 1.5A**). For example, in a symmetric bimolecular structure having lateral loops, the G-tetrads are formed from guanosines of alternating *anti-syn-anti-syn* conformation, with each *syn*-guanosine donating hydrogen bonds to an adjacent *anti*-guanosine, and accepting hydrogen bonds from the other adjacent guanosine [32]. This results in a rectangular G-tetrad core with grooves of alternating wide-narrow-wide-narrow widths (**Fig. 1.5B**). In contrast, a bimolecular quadruplex with diagonal loops, has guanosines that adopt the *syn-syn-anti-anti* conformations, which results in G-tetrads that adopt a parallelogram arrangement and the formation of alternative wide, medium, narrow, medium width grooves between strands (**Fig. 1.5C**).

Groove dimensions are then variable, and depend on overall topology and the nature of the loops. Grooves in quadruplexes with only lateral or diagonal loops are structurally simple, but, in contrast, grooves that incorporate propeller loops have more complex structural features that reflect the insertion of the variable-sequence loops into the grooves.
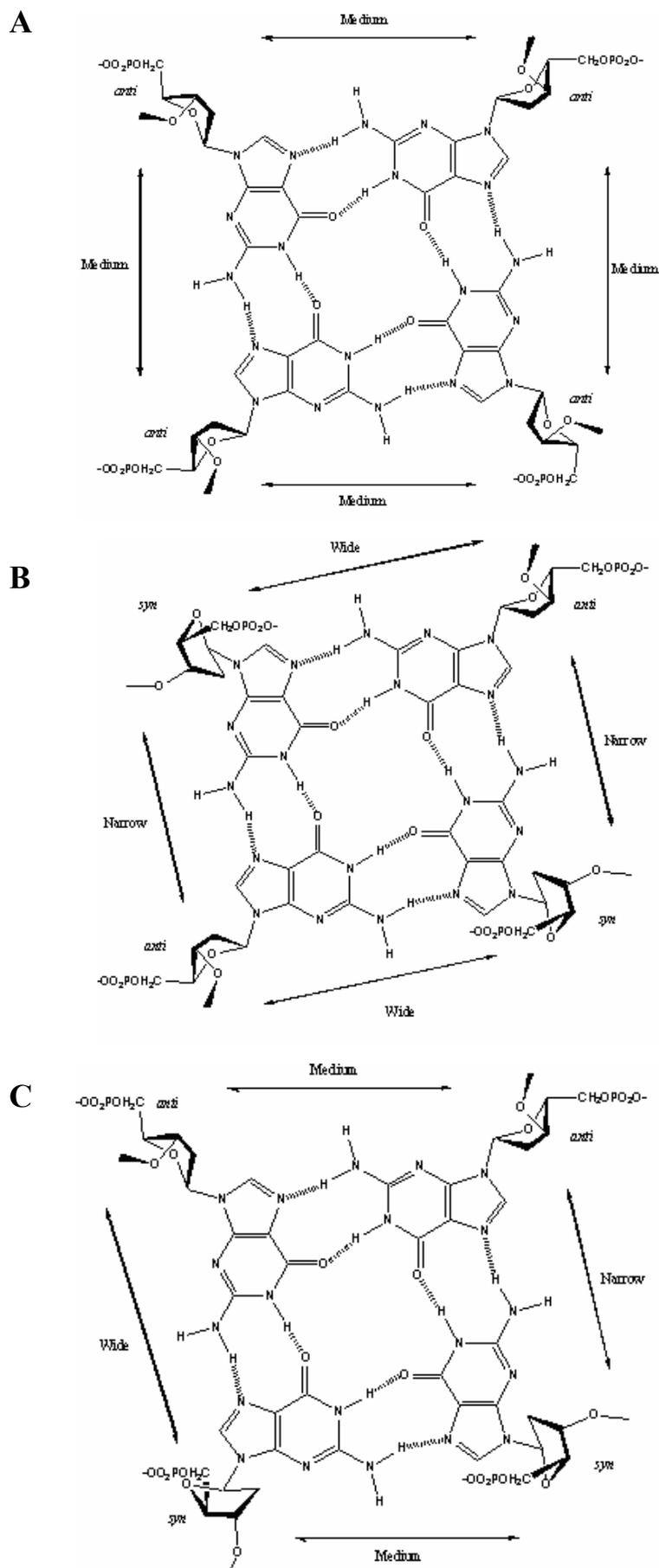
Fig. 1.5 – *Variations in glycosidic torsional angles and their effects on groove width in G-quadruplex structures. All anti (A), anti-syn-anti-syn (B), and syn-syn-anti-anti (C) G-tetrads.*

1.3.3 G-Quadruplex metal ion binding

When one or more G-tetrads are stacked, a cylindrical central cavity is produced [33]. This cavity, lined with the guanine O6 carbonyl oxygens, forms a specific binding site for metal ions [34]. The coordination of potassium [35], sodium [36], and strontium [37] all provide both thermodynamic and kinetic stability to the G-quadruplex structure. Experiments demonstrate that the G-quadruplex is stabilized by the alkali series in the following order: $K^+>>Na^+>Rb^+>NH_4^+>Cs^+>>Li^+$, and for the alkaline earth series in the order: $Sr^{2+}>>Ba^{2+}>Ca^{2+}>Mg^{2+}$ indicating that the atomic radii of 1.3Å of potassium and strontium fit best in the coordination site between adjacent G-tetrads [38, 39]. The precise location of the cations between the tetrads is dependent on the nature of the ion. Na^+ ions within the channel have been observed in a range of geometries; in some structures, a Na^+ ion is in plane with a G-tetrad whereas in others it is between two successive G-tetrads [17]. K^+ ions are always equidistant between each G-tetrad plane, interacting equally with each of eight carbonyl oxygens in a bipyramidyl antiprismatic coordination geometry [40].

The selectivity of G-quadruplex DNA for K^+ versus Na^+ ion has been studied by Feigon and coworkers by ^{1}H NMR, using the bimolecular quadruplex formed by $d(G_3T_4G_3)$ as a model system [41]. In this system, the G-quadruplex binds two K^+ or Na^+ ions, corresponding to one metal ion sandwiched between each pair of G-tetrads. Using competition NMR experiments, Feigon and co-workers determined that there is only a modest difference in free energy of 1.7 kcal/mol favoring the binding of K^+ versus Na^+ ions. They suggest that this modest free energy difference is a result of the contributions of the relative free energies of hydration, which favors K^+ binding [41].

A biologically significant aspect of metal ion binding is related to the observation of different conformation in the presence of different cations. There are a number of well-established examples where the change from Na^+ to K^+ induces profound structural alteration, implying high conformational flexibility for these particular quadruplexes. For example, studies on the DNA sequence $[d(GGCT_4GGGC)]_2$ in the presence of either K^+ or Na^+ show distinct differences. In the presence of K^+ ion, two G-tetrads are stacked in an antiparallel form bimolecular quadruplex [42]. The smaller Na^+ ion, however, allows for the formation of non-Watson-Crick G:C base pairs in the connecting loop region. These GCGC stabilized tetrads stack on the flanking ends of the G-tetrads with Na^+ ions coordinated between the G-tetrad and its neighboring GCGC tetrad [43].

On the other hand, it is equally clear that some quadruplexes, have very stable and unique topologies. For example, the X-ray structure of the bimolecular quadruplex formed by the *Oxytricha nova* sequence $d(G_4T_4G_4)$ in the presence of K^+ ions [27], shows identical quadruplex topology to the NMR structures in Na^+-containing solution [24-26], which is itself identical with the NMR structure in Tl^+-containing solution [44].

1.4 Biological interest of G-quadruplexes

Recently, there has been growing interest in the study of quadruplexes because of their possible involvement in many biological processes. Telomeric DNA in a variety of eukaryotic organisms is organized in tandem repeats of short G-rich sequences [45-51] (**Table I**) that were shown to form G-quadruplex structures [22, 31, 52, 53]. The intramolecular G-quadruplex formation observed in the human telomeric sequence d[TTAGGG]$_4$ can inhibit telomerase activity [54-56]. As *in vitro* telomerase induction can transform healthy cells into malignant ones [57], telomerase inhibition by targeting G-quadruplex leads to telomere shortening and senescence in tumor cells [58-67]. G-quadruplex structures have been shown to exist *in vivo* in *Stylonychia lemnae* macronuclei [68] and recently in human cells [69].

Table I - Sequences and length of Telomeric repeats

Organism	Repeat	Length
Oxytrichia	TTTTGGGG	20 bp
Euplotes	TTTTGGGG	28 bp
S. cerevisiae	TG$_{2-3}$(TG)$_{1-6}$	300 bp
Tetrahymena	TTGGGG	300-400 bp
Arabidopsis	TTTGGGG	2-15 kbp
Mus spretus	TTAGGG	5-15 kbp
Homo sapiens	TTAGGG	5-25 kbp
Mus musculus	TTAGGG	20-50 kbp

Quadruplex-forming G-rich sequences are also found in a number of cancer-related genes such as *c-myc* [70-73], *c-kit* [74, 75], *HIF-1α* [76], *VEGF* [77], *BCL-2* [78, 79], *KRAS* [80], *Rb* [81, 82] and *RET* protooncogene [83]. The evidence for the involvement of G-quadruplexes in transcriptional control of the *c-myc* oncogene has been shown [71]. The overexpression of this oncogene is associated with a variety of human malignancies, including breast, colon, cervix, small-cell lung carcinomas, osteosarcomas, glioblastomas, and myeloid leukemias [84-87]. In addition, c-Myc protein can induce telomerase activity by increasing the transcription rate of hTERT [88]. The nuclease hypersensitivity element (NHE) III$_1$ upstream of the P1 promoter of *c-myc* controls up to 90% of the transcriptional activation of this gene [71, 89-93]. The element NHE III$_1$ can form transcriptionally active and silenced forms [94]. The guanine-rich segment of this element is part of the silenced form and can adopt a quadruplex structure that is critical for transcriptional silencing [71, 73]. G-quadruplex interacting agents with specificity for binding to silencer element can be good anticancer agents [95].

G-quadruplexes are also important because they have been implemented in the design of novel aptamers aimed at binding and inhibiting particular proteins [96-98]. For example, the sequence d(GGTTGGTGTGGTTGG) (TBA: thrombin binding aptamer) was found to be a potent inhibitor of thrombin in a fibrinogen clotting assay [99, 100]. Furthermore, quadruplex forming oligonucleotides have resulted to be potent inhibitors of the HIV-1 integrase, the enzyme responsible for the insertion of viral DNA into the host genome [101]. Recently, several non-telomeric or telomeric G-tail oligonucleotides have been found to exhibit antiproliferative activity against many tumor cells in culture [102].

1.5 Investigated systems and methodologies employed

Understanding the energetic processes that dictate the structure, stability, and affinity of G-quadruplexes has achieved new importance in the last years, with the recognition that minute alterations in these properties can dramatically alter the function of the macromolecule, often with severe consequences. Moreover, the successful of the rational design of drugs that selectively bind to a G-quadruplex structure, requires that the thermodynamic effect of a chemical alteration be predicted with considerable confidence.

In this context, the general purpose of this thesis was to study the energetic aspects of G-quadruplex structures and of their interaction with proteins and potential anticancer agents.

In the first part of the thesis, isothermal titration calorimetry technique has been applied to understand the thermodynamic proprieties of drug-quadruplex interactions and for screening among various drugs. The quadruplexes taken into consideration were the unimolecular and the bimolecular human telomeric quadruplexes, of sequence d[AGGG(TTAGGG)$_3$] and d(TAGGGTTAGGG), respectively; and the tetramolecular quadruplex formed by the truncated sequence of *Oxytricha* and *Tetrahymena* telomeric DNA, d(TGGGGT). The quadruplex-binding ligands investigated were the distamycin A (Dist-A), a small molecule which binds with high affinity to the minor groove of B-form DNA [103] that, recently, has been shown to interact also with quadruplex DNA [104, 105]; the compounds **1** and **2**, two derivatives of distamycin; and the cationic porphyrin TMPyP4 (5,10,15,20-tetrakis-(*N*-methyl-4-pyridyl)-21*H*,23*H*-porphyrin).

In the second part of the thesis a combination of calorimetric and computational methodologies have been used to characterize the physico-chemical and the binding proprieties of a modified thrombin binding aptamer (mTBA), whose oligonucleotide sequence, containing a 5'-5' site of polarity inversion, $^{3'}$GGT$^{5'}$-$^{5'}$TGGTGTGGTTGG$^{3'}$ adopts an intramolecular G-quadruplex structure in solution [106]. This study was performed comparing the mTBA with its unmodified counterpart, the quadruplex-forming aptamer TBA.

In the last part of the thesis molecular dynamics simulations have been utilised to study the differences between two DNA and RNA 14-mer quadruplexes of analogous sequences. In particular, the investigated molecules were the bimolecular quadruplex from the sequence d(GGAGGTTTTGGAGG) that was found to form a quadruplex with parallel and antiparallel strand orientations composed by four canonical G-tetrads and two T4 diagonal loops [107]; and the corresponding RNA sequence r(GGAGGUUUUGGAGG) that, in contrast, folds into an unusual quadruplex, comprising a G-tetrad, a U4 double-chain reversal loop and GGAGGA hexad, the interfaces of which promote a dimeric structure [108].

The physico-chemical characterization of these systems was therefore principally performed by means of calorimetric methodologies as isothermal titration calorimetry and differential scanning calorimetry, and by computational techniques as docking and molecular dynamics simulations.

CHAPTER 2: METHODS

2.1 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is a valuable tool for characterizing interactions of G-quadruplexes with other biomolecules, including small ligands, thanks to its general applicability and precision [109, 110]. ITC is a high-accuracy method for measuring binding affinities and stoichiometry; moreover, it is the only technique that directly measures the binding enthalpy. ITC also allows dissecting the free energy of binding into enthalpic and entropic components to reveal the overall nature of the forces that drive the binding reaction [111, 112].

In ITC, one component of a complex (such as DNA or a protein) is present in the calorimeter's sample cell, and the second component (for example, a drug) is slowly added in an incremental, stepwise fashion. Since all binding events are accompanied by the evolution or absorption of heat (a change in enthalpy), the analysis of these extremely small thermal effects arising from the binding allows a full thermodynamic characterization of the reaction and provides fundamental information about the molecular interactions driving the process.

The calorimeter used for our experiments is the CSC 4200 ITC. A schematic diagram of the ITC is shown in **Fig. 2.1**. The calorimeter holds two removable vessels (reference and sample cells), with a nominal volume of 1.3 ml, that are contained in an ultra-stable constant temperature bath. The reference vessel contains water or buffer solution, the sample vessel holds a solution containing one of the reactants. The two vessels are constantly kept in thermal equilibrium with the bath during the experiment. The heat flow between the reaction vessel and the isothermal block is precisely measured by thermoelectric device sensors that surround the vessel and is monitored as a function of time. The ITC measures the heat generated/absorbed when two solutions of reactants are mixed. The other reactant is contained in a precision syringe (25-250 μl) connected to the sample cell and is added in computer-controlled injections (1-20 μl). The solution in the sample vessel is stirred with a propeller to ensure rapid mixing of components. The heat change caused by any reaction occurring when the solutions are mixed, is detected by the thermoelectric device sensors, amplified and converted by an electronic output circuit to a signal corresponding to the heat change. The signal is registered by the instrument as a deflection peak. The CSC 4200 ITC can detect heat effects as small as 0.4 milliJoule, allowing titrations to be done with nanomoles of biopolymer, moreover, ITC may be run at any temperature between 0 and 110°C.

In a typical ligand-macromolecule titration, the chemical reaction generated by each injection either releases or absorbs a certain amount of heat (q_i) proportional to the amount of ligand that binds to

the macromolecule in a particular injection (V×ΔL$_i$) and the characteristic binding enthalpy (ΔH) for the reaction:

$$q_i = V \times \Delta H \times \Delta L_i$$

where V is the volume of the reaction cell and ΔL$_i$ is the increase in the concentration of bound ligand after the ith injection [113]. The heat after each injection is therefore obtained by calculating the area under each peak. Because the amount of uncomplexed macromolecule available progressively decreases after each successive injection, the magnitude of the peaks becomes progressively smaller until complete saturation is achieved. Once this situation is reached, subsequent injections produce similar peaks corresponding to dilution or mechanical effects that need to be subtracted from all the injection peaks before analysis.
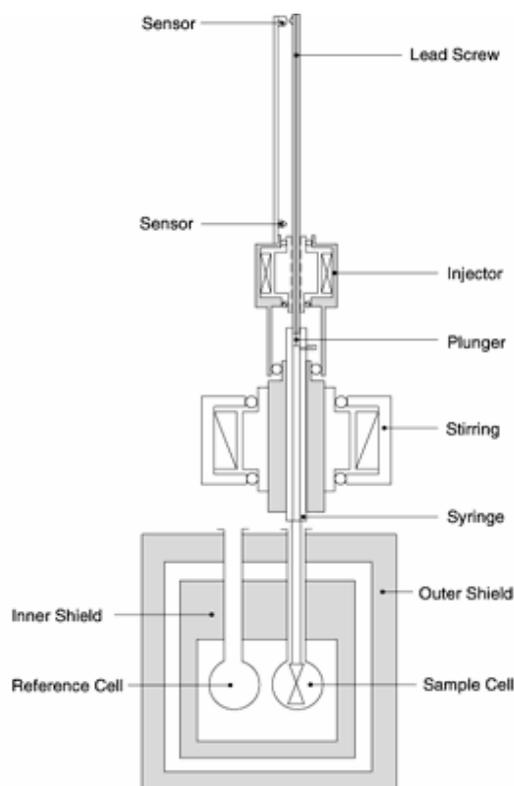
Fig. 2.1 - *Schematic representation of an ITC.*

The top panel of **Fig. 2.2** depicts the signal produced by the sequence of injections, and the bottom panel shows the data after integration of each injection peak. The sigmoidal shape of the bottom panel (**Fig. 2.2**), with numerous data points throughout the curved rise portion of the plot, facilitates estimation of the midpoint of the transition, and thus the stoichiometry of the binding reaction (in

this case, 1:1). K_b and ΔH are calculated by iterative approximation. A value for K_b is initially estimated, then the concentration of bound complex is calculated for each injection. In combination with the measured heat, these values are used to determine the average of ΔH. The ΔH and the calculated concentration are then used to determine an expected heat per injection, and the error square sum between the measured and expected heat for each peak is calculated. The value of K_b is then adjusted and the process repeated until a minimum error square sum is obtained.

The values of K_b, ΔH and n (stoichiometry) can often be calculated from a single experiment as long as the concentration of both macromolecule and ligand are accurately known [114] and chosen so that:

$$10 < K_b\,[M]T < 1000$$

where [M]T is the total concentration of macromolecule in the sample cell titrated by ligand [115]. If concentrations are not within this range, the curvature of the titration plot can be so low as to be almost linear, or so high as to produce a step-like profile. In these cases K_b may not be estimated accurately. Typically, macromolecule concentrations in the order of 10-100 μM are used, permitting K_b values in the range $10^2 - 10^9$ M^{-1} to be accurately estimated.

Since temperature (T) is held constant throughout the entire experiment, the free energy ($\Delta G°$) of the binding reaction can be determined by:

$$\Delta G° = -\,R\,T\,\ln K_b$$

where R is the gas constant.

ITC directly measures $\Delta H°$, so the change in entropy ($\Delta S°$) can be determined by:

$$\Delta S° = (\Delta H° - \Delta G°) / T$$

Quantification of these thermodynamic parameters reveals the physical processes involved in the binding reaction. A spontaneous binding process must have a negative $\Delta G°$, and $\Delta G°$ will become increasingly negative as binding becomes tighter. As seen above, free energy changes have both an enthalpic and entropic component. The enthalpic contribution to binding is primarily due to an increased number of hydrogen bonds at the ligand-target interface, and to more favorable van der Waals interactions between the two interacting molecules; the hydrophilicity of the system will determine how important electrostatic, polar and dipolar interactions will be in driving the reaction.

The entropic contribution has two primary components: conformational changes, such as folding or unfolding of the macromolecules, and the release of bound solvent as hydrophobic groups interact. The large number of ordered water molecules released into the bulk solvent when the hydrophobic surfaces of the ligand and target interact provides the main driving force for hydrophobic interactions. This driving force is sufficient to compensate for the unfavourable conformational entropy of the macromolecule and ligand caused by decreased conformational and rotational freedom following binding. In addition to the entropic effect, burial of surface area also affects the heat capacity of the sample, since water molecules ordered at hydrophobic surfaces have a different heat capacity from that of water that has been released into the bulk solvent following binding.
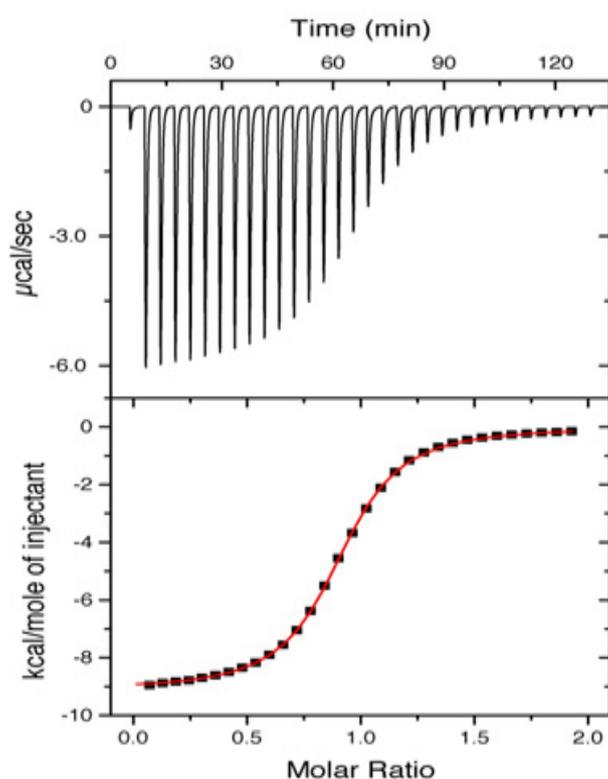

Fig. 2.2 – *(top panel) Signal produced by the sequence of injections for a tipical ITC experiment. (bottom panel) Data after integration of each injection peak.*

2.2 Differential scanning calorimetry

Differential scanning calorimetry (DSC) is the most direct and sensitive approach for characterizing the energetics of conformational transitions of biological macromolecules, such as G-quadruplexes [116]. DSC is a useful technique because it leads to determine, in a single measurement, the temperature, the enthalpy and the heat capacity changes associated to the transition process. In particular, the enthalpy, being a thermodynamic potential, contains information on the states to which the system can belong within the investigated temperature range.

The principle of scanning microcalorimetry is to measure enthalpy changes of a solution of the macromolecules as a function of temperature change. This is obtained by measuring the power required to keep a sample at the same temperature as that of a reference solution as the temperature of both is increased in a linear manner.

Biological and pharmaceutical products are usually studied in diluted solutions. Consequently, the associated thermal effects are very low, so that for detecting these phenomena highly sensitive DSCs, called micro DSC, can be used. The Setaram micro DSC has been used in our laboratory. The calorimetric block of the micro DSC consists of a gold-plated metal cylinder with high thermal conductivity (**Fig. 2.3**). This assembly ensures excellent temperature uniformity with accurate control, which in turn determines the very good baseline stability in the calorimeter. Two hollows are machined in the block to accommodate the experimental vessel. The micro DSC is equipped with two closed steel vessels (850 μL) that are removable and can be easily cleaned. The "measurement" vessel holds the sample to be analyzed. The "reference" vessel contains buffer to compensate for the thermal effect associated with heating of the samples. Various types of experimental vessels can be used depending on the applications. The micro DSC features a three-dimensional sensor with Joule effect calibration for highly sensitive and precise calorimetric measurements. Each vessel is surrounded by very high sensitivity Peltier elements ensuring the thermal link with the calorimetric block. These detectors are good thermal conductors that keep the temperature in the vessels identical to that in the calorimetric block. The heat-flow transducer enables the micro DSC to reach a very high sensitivity detection limit of 0.2 mW with very good measurement accuracy. The temperature of the calorimeter is set by thermostatic control of a circulating liquid which is heated or cooled by the controller from $-20°C$ to $+120°C$. External water circulation removes the heat. The transducer zone is protected by thermal buffers that prevent outside interference but provide direct access to the vessels. Using a liquid thermostat and precisely controlling the temperature reduces the effect of outside temperature fluctuations and allows the accurate measurement of very small thermal effects. The calorimeter was interfaced to an IBM PC computer for automatic data collection and analysis using previously described software [117].
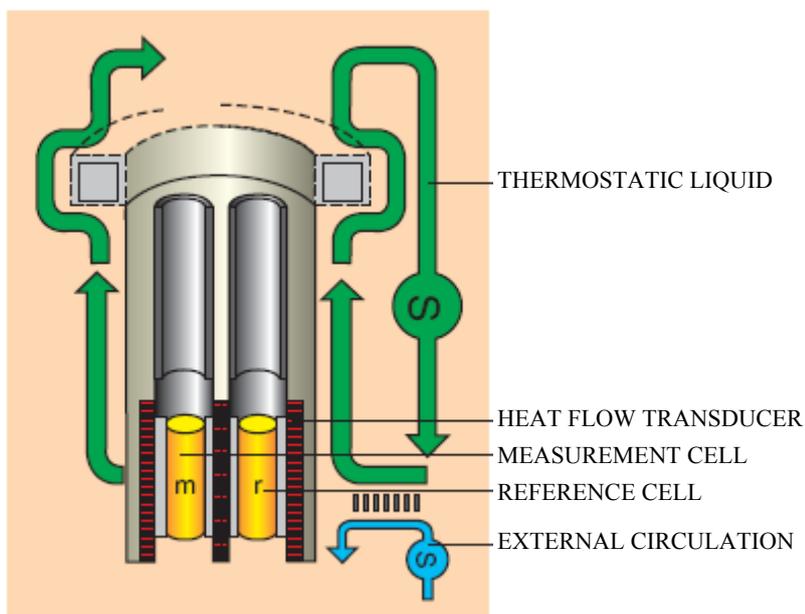
Fig. 2.3 - *Calorimetric block.*

For biological solutions, the sample normally contains the macromolecule in buffer and the reference is the buffer solution. As the temperature changes, the macromolecule undergoes a transition in a cooperative fashion. This transition (protein denaturation or DNA dissociation) arises from the destruction of the numerous small forces that stabilize the native structure. Such disruption changes the enthalpy of the system giving rice to a drop in temperature, because the process is usually endothermic. The calorimeter will provide energy to the sample to maintain its temperature at the same value as that of the reference solution. The energy is measured as a power input and the raw output is a data set of power vs. temperature. Power is easily converted to the apparent molar excess heat capacity using the following equation:

$$< \Delta C_P^0 >= \frac{P}{\sigma m}$$

where $< \Delta C_P^0 >$ is the apparent excess heat capacity in J mol^{-1}K^{-1}, P is power in J s^{-1}, σ is the scan rate in Ks^{-1} and *m* is the number of moles of the protein in the sample. Subtraction of the buffer baseline corrects the data for the partial molar heat capacity of the solvent, allowing the partial molar heat capacity of the macromolecule to be determined. Thus, a typical calorimetric curve reports the excess heat capacity function, $< \Delta C_p >$, versus temperature (**Fig. 2.4**). The experimental transition enthalpy, $\Delta H°(T_m)$, is obtained by integrating the area under the curve. The transition

midpoint (T_m, or the melting temperature) is the temperature corresponding to the maximum of DSC peak. ΔC_P^0 is the overall denaturation heat capacity, i.e. the difference between the denatured and native macromolecule heat capacity.

The $\Delta H°(T_m)$ is actually a net value from a combination of endothermic contributions, such as the disruption of hydrogen bonds, and exothermic processes such as the disruption of hydrophobic interactions. The sharpness of the transition peak is indicative of the cooperative nature of the unfolding process. If the unfolding produces a narrow, symmetric peak such as in **Fig. 2.4**, the transition is close to a two-state, reversible and highly cooperative process [118]. However, it is frequently convenient to compare the directly measured enthalpy with indirect estimates using the classical van't Hoff equation:

$$d(\ln K(T))/dT = \Delta H_{v.H.}/RT^2$$

The van't Hoff analysis is based on a hypothesis of model of the process involved. In a two-state thermal denaturation process, the $\Delta H_{v.H.}$ is the enthalpy change per mol of cooperative units, as defined by the model. In this case, the ratio between the values of $\Delta H°(T_m)$ and $\Delta H_{v.H}$ is close to unity. If the unfolding transition is not two-state, but involves one or more intermediates, the transition appear broader then expected and $\Delta H_{v.H}$ is less than $\Delta H°(T_m)$. If the macromolecule unfolds cooperatively as a dimer or higher oligomers, the transition appears sharper than expected for a two-state transition of a monomer and the $\Delta H_{v.H}$ is greater than $\Delta H°(T_m)$.

In order to obtain correct thermodynamic data, the thermal transition must be fully reversible. This is verified by using the reheating criterion: reheating the sample after cooling and obtaining a second calorimetric curve superimposable on the first.
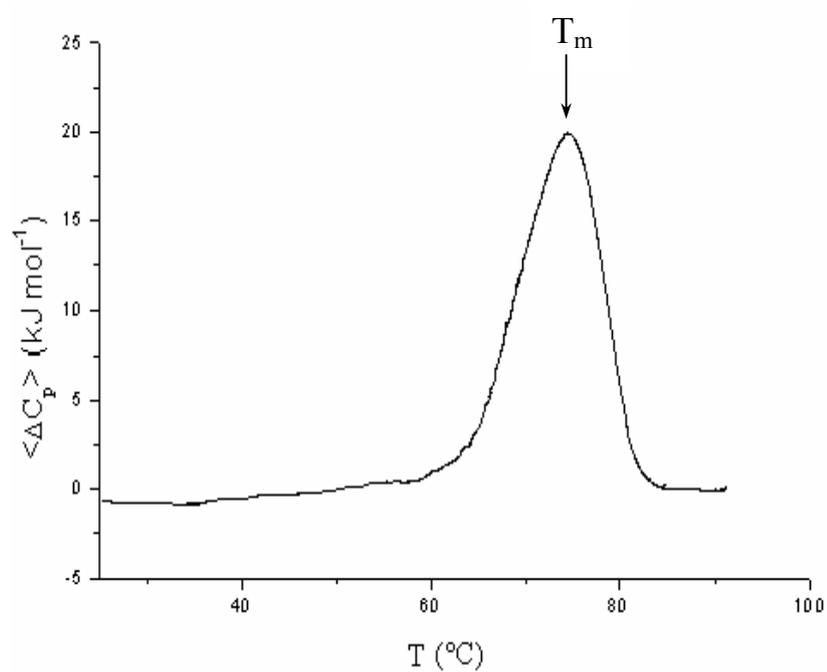
Fig. 2.4 – *Example of calorimetric curve for a duplex DNA dissociation. The arrow indicates the maximum of DSC peak corresponding to the melting temperature.*

2.3 Docking

Docking is the process by which two molecules fit together in 3D space. This methodology is often used to predict protein-protein, protein-DNA (or RNA), and protein-small molecule interactions. Predictive docking procedures start from the individually determined (unbound) structures of two molecules and they aim at predicting the complex structure. All docking programs contain a scoring function to discriminate between near-native docked orientations and incorrect orientations, and a search algorithm that is designed to sample rapidly all possible docking orientations. Proteins undergo conformational changes upon complex formation to varying extent. Even though structural flexibility is mostly restricted to surface side chains [119], the innate flexibility renders docking extremely difficult. The algorithm by Abagyan and colleagues [120, 121] allows for surface side chain flexibility, however, such algorithms typically require hundreds of hours for each complex [120]. The alternative is to adopt the *rigid-body* approach, which samples only the six rotational/translational degrees of freedom, but uses target functions that are tolerant to some overlap of the two molecules being matched. ZDOCK (rigid-body docking) [122] is a accurate program with a simple target function. In particular, it is essential to have some correct predictions among the complexes that are produced. The terms employed during initial stage docking are shape complementarity, desolvation free energy and electrostatic energy [123]. The ZDOCK program uses two PDB files as input (the two molecules to dock) and creates a list of top-scoring ligand orientations as output. For every rotational orientation, ZDOCK keeps the 10 best translational orientations. The basic search algorithm samples exhaustively the entire rotational and translational space of the ligand with respect to the receptor, which remains fixed at the origin. For every rotation, the algorithm rapidly scans the translational space using fast Fourier transformation (FFT). Three lists are kept at all times, each containing ligand orientations: the first list with the best shape complementarity score S_{SC}, the second list with the best shape complementarity and desolvation score $\alpha S_{SC} + S_{DS}$, and a third list with the best shape complementarity, desolvation, and electrostatics score $\alpha S_{SC} + S_{DS} + \beta S_{ELEC}$.

Shape complementarity (SC) is the most basic ingredient of all scoring functions for docking. As the name implies, it is a geometric descriptor, stemming from the practical observation that molecules surfaces are complementary to each other at the binding interface.

The total Desolvation Free Energy (DS) of complex formation (S_{DS}; DS stands for desolvation) is calculated by summing the Atomic Contact Energy (ACE) scores of all atom pairs between the receptor and the ligand that are within 6 Å distance [123]. ACE is a desolvation free energy measure, defined as the free energy of replacing, for example, a protein-atom/water contact with a protein-atom/protein-atom contact. Combined with an electrostatics term, ACE has been shown to

accurately predict the free energy changes associated with transferring amino acids from the protein interior to water, to predict site-specific mutations of hydrophobic amino acids and to assess protease-inhibitor binding [124-127].

The electrostatic energy (ELEC) can be expressed as a correlation between the electric potential generated by the receptor and the charges of the ligand atoms, as described by the Coulomb formula [128]. Additionally, grid points in the core of the receptor are assigned a value of 0 for the electric potential to eliminate contributions from non-physical receptor-core/ligand contacts.

However, in an unbound docking study, surface side chains tend to be at non-optimal positions and therefore the resulting electrostatic energy can be inaccurate and even unrealistic. Artifacts can also be introduced by the grid representation. Therefore the electrostatics energy S_{ELEC} is scaled by a factor β, and the shape complementarity score is weighted by a scaling factor α. The final target function is:

$$S = \alpha S_{SC} + S_{DS} + \beta S_{ELEC}$$

The default values for scaling factors are α= 0.01 and β= 0.06.

One useful feature of ZDOCK is that it allows the user to "block" residues, making it unfavorable for selected residues to be in the interface. This can be useful when, for example, experimental and/or biological information is known regarding the location of the binding site. Blocking is accomplished by assigning a special atom type to all atoms of the residues to be prohibited from the putative binding site.

Protein docking methods have improved substantially over the past few years, however, since DNA and RNA are highly charged and more flexible than proteins, predicting protein–nucleic acid interactions by rigid body docking is more difficult compared to protein–protein interactions. For this reason, a very limited number of structure predictions of protein–nucleic acid complexes have been reported so far [129-134].

2.4 Molecular Dynamics methods

Molecular Dynamics (MD) is a computer simulation technique in which the time evolution of a set of interacting atoms is computed by integrating their equations of motion. Therefore, MD allows us to view how a molecular system evolves through time and to derive average properties of the system, given a simulation of sufficient length. MD is particularly useful when the system cannot be studied by experimental methods.

The equations of motion can only be solved numerically for a multi-body problem. To calculate the dynamics of the system (*i.e.* the position of each atom as a function of time), Newton's classical equations of motion are solved for each atom given an empirical force field:

$$F_i = m_i \, a_i \tag{1.1}$$

where F_i is the force exerted on particle i, m_i is the mass of particle i and a_i is the acceleration of particle i. The force on each atom is the negative of the derivative of the potential energy (V) with respect to the position of the atom ($r_1, r_2,\ldots, r_N$):

$$F_i = -\frac{\partial V}{\partial r_i} \tag{1.2}$$

Once the coordinates of the atoms of a starting structure and their velocities are defined, the force acting on each atom can be calculated for each point in time t + dt and a new set of coordinates can be generated. The repetition of this procedure generates a molecular trajectory corresponding to the time-dependent fluctuations of the atomic positions. The accuracy of the simulations is directly related to the potential energy function that is used to describe the interactions between particles. In MD a classical potential energy function is used that is defined as a function of the coordinates of each of the atoms.

The potential energy function is separated into terms representing covalent interactions and non covalent interactions. The covalent interactions may be described by the following terms:

$$V_{bond} = \sum_{i=1}^{N_b} \frac{1}{2} k_i^b (r_i - r_{0,i})^2 \tag{1.3}$$

$$V_{angle} = \sum_{i=1}^{N_\theta} \frac{1}{2} k_i^\theta (\theta_i - \theta_{0,i})^2 \tag{1.4}$$

$$V_{dihedral} = \sum_{i=1}^{N_\phi} \frac{1}{2} k_i^\phi \cos(n_i(\phi_i - \phi_{0,i})) \tag{1.5}$$

$$V_{improper} = \sum_{i=1}^{N_\xi} \frac{1}{2} k_i^\xi (\xi_i - \xi_{0,i})^2 \tag{1.6}$$

the equations correspond to two, three, four and four body interactions, respectively. These interactions are represented by harmonic potentials for the bond lengths r_i, for the bond angles θ_i, for the improper dihedral (out of the plane) angle ξ_i and by a more complex potential for the dihedral angles Φ_i. The non-covalent (non-bonded) interactions, which correspond to interactions between particles separated by more than three covalent bonds, are usually described by Coulomb's law:

$$V_{Coulomb} \quad = \quad \sum_{i<j} \frac{1}{4\pi\varepsilon_0\varepsilon_r} \frac{q_i q_j}{r_{ij}}$$

(1.7)

for the electrostatic interactions (for a pair of atoms carrying the partial charges qi and qj), and by a Lennard-Jones potential:

$$V_{LJ} \quad = \quad \sum_{i<j} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}$$

(1.8)

for the van der Waals (vdW) interactions, where rij is the atomic distance between particles i and j. The force field parameters describe the strength of the interactions. For bonded interactions parameters are defined for bond stretching, bond bending and torsional rotation. Another set of parameters determines the strength of non-bonded electrostatic and van der Waals interactions. Electrostatic interactions are generally represented by point charges located at the center of the atom.

It should be kept in mind, however, that MD is affected by several limitations [135]. Firstly, MD is computationally very demanding and the computational load scales with the square of the system size. Simulation times are currently limited to hundreds of nanoseconds or a few microseconds at most. The phenomena that can be explored must occur with sufficient statistical significance within time scales that are encompassed by the computation.

How the system evolves through time is specified by the force field and by an integration time step that determines where the atoms will be positioned at time t + dt. MD requires the use of a very small time-step (1-2 fs) to achieve accurate results, because small time-steps limit the approximations that are introduced by the numerical integrator. This limits the overall scope of the simulated time and the computable properties.

According to statistical thermodynamics, physical quantities are represented by averages over configurations belonging to a certain statistical ensemble. A trajectory obtained by molecular dynamics provides such an ensemble. Therefore, a measurement of a physical quantity by simulation is simply obtained as an arithmetic average of the various instantaneous values adopted

by that quantity during the MD run. Statistical thermodynamics is the link between the microscopic ensembles and and the macroscopic properties. In the limit of an exact force field and very long simulation times, one could expect the phase space to be fully sampled and in that limit the averaging process would yield exact thermodynamic properties. In practice, MD runs are always of finite length and one should exert caution when judging the sampling quality. An important constraint in deriving average properties is to extract configurations only from an ensemble at thermal equilibrium. Therefore, MD simulations start generally with an equilibration phase. Once equilibrium is reached, the simulation enters the production phase. The production run should be long enough to sample the property of interest with sufficient statistical significance.

CHAPTER 3: THERMODYNAMIC STUDY OF QUADRUPLEX-DRUG INTERACTION

3.1 Telomeres and telomerase

In the last years telomeres and telomerase have been the topic of thousands of papers due to their role in tumourigenesis. Telomeres are DNA-protein complexes which preserve chromosomal ends from degradation, recombination or end-to-end fusion [136-139]. They have also a variety of other functions, including role in chromatin organization and cell proliferation [140, 141].

Telomeric DNA of eukaryotes consists of repetitive G-rich sequences highly conserved during evolution (see **Table I**), organized in double strand for most of its length and in single strand at the 3'-ends. In particular, human telomeric DNA contains 5-25 kb pairs of the repetitive sequence TTAGGG and terminates with 3'single-stranded G-rich overhangs [46, 142, 143]. The G-rich overhangs of telomeric DNA are prone to fold back into G-quadruplex structures.

Electron microscopic studies have suggested that telomeres are not a linear structures, as believed in the past, but form loop structures with the 3'-G-rich strand, called t-loop (telomere loop) intercalating the duplex telomeric repeats of the 5'-end to form the D-loop (displacement loop) (**Fig. 3.1**) [136]. Telomere-specific proteins bind to the single- and double-stranded telomeric DNA, and contribute to stabilize these loops and to protect chromosome ends. In particular, a proteins complex named shelterin has been recently identified (**Fig. 3.1**) [144]. Shelterin is formed by six components that have been gradually identified over the past 15 years: TRF1, TRF2, TIN2, Rap1, TPP1 and POT1 [137, 145-150]. Three components, TRF1, TRF2 (Telomeric Repeat Binding Factor 1 and 2), and POT1 (Protection Of Telomere 1), directly recognize TTAGGG repeats. These proteins are interconnected by three other proteins: TIN2 (TRF1-interacting nuclear factor 2), Rap1 (Repressor/activator protein 1) and TPP1(Telomerase associated protein 1). The most conserved component of shelterin, POT1 binds directly to the 3'-overhang and through interactions with the TRF1-duplex complex [151, 152], whereas TIN2 and TPP1 are considered key components in mediating the six-protein complex assembly [153]. There are also other proteins at chromosome ends that are not part of shelterin [154] that, together with the shelterin complex, have been proposed to provide the basis for constructing an interaction map of telomere regulators in mammalian cells [155].

Telomerase is the enzyme implied in maintaining the telomeric DNA length at the ends of chromosomes. When the telomerase is not operative, chromosomes are progressively eroded after each cell replication [59] until a critical state of replicative senescence is reached, with no further replication, and cell death occurs [156, 157]. Telomerase is active in about 85% of cancer cells, thus

favouring their immortalization [158, 159]; it is inactive in normal somatic cells, and active in germ and stem cells with regenerative potential [160].
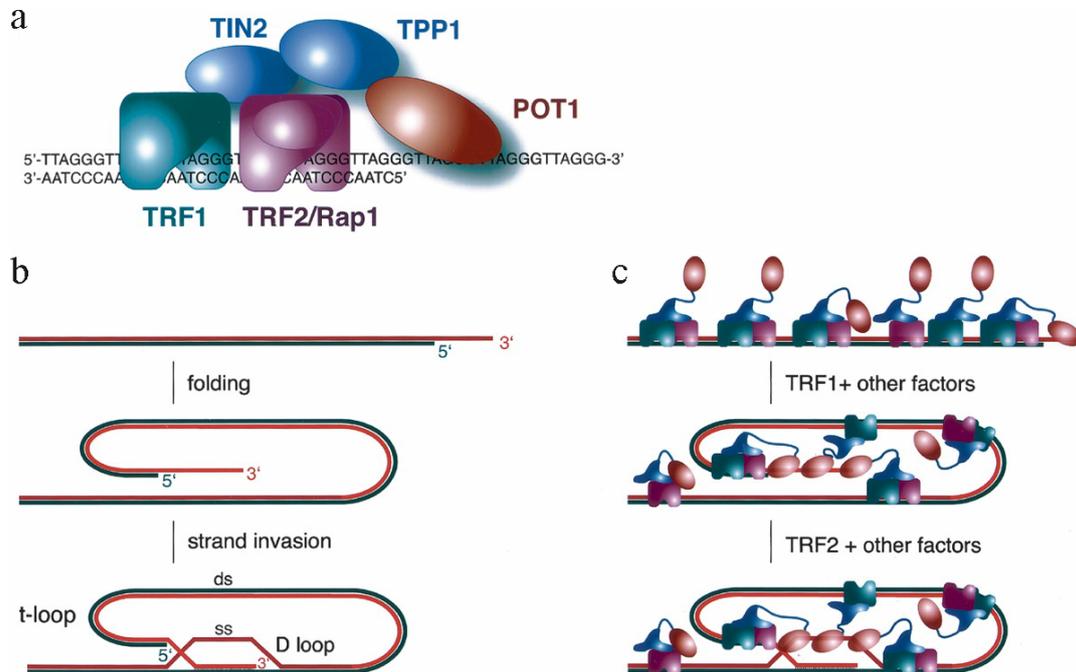
Fig. 3.1 - *(a) Schematic representation of the six known subunits of shelterin on telomeric DNA. (b) The t-loop structure. The 3'-overhang invades the adjacent duplex telomeric repeat array, forming a D-loop. (c) Speculative model for t-loop formation by shelterin. TRF1 could potentially fold the telomere, whereas TRF2 can mediate t-loop formation in vitro.*

Human telomerase is a ribonucleoprotein complex containing two major components as well as many associated proteins that are critical for function. The first major component is the hTERT (human Telomerase Reverse Transcriptase), a reverse transcriptase catalytic subunit; the second component is the hTR (human Telomerase RNA), an RNA subunit [161, 162]. hTERT is a 120 kDa protein, active as a dimer [163], that interacts with telomeric DNA and telomere-binding protein [154, 164-168]. The interaction of hTR with hTERT determines the catalytic activity [159, 169]. Telomerase uses its internal RNA component, complementary to the telomeric single-stranded overhang, as a template for the synthesis of telomeric TTAGGG repeats directly onto the ends of chromosomes. The enzyme adds the first six bases, then translocates the template RNA for the synthesis of the next six bases (**Fig. 3.2**) [170]. This extension at the 3′-end of the DNA template compensates for telomere shortening that results from nuclease action and incomplete terminal DNA replication. A multi component "telomere homeostasis" system promotes this telomeric

extension when a telomere becomes shortened, thus making telomeric repeats accessible to the telomerase [164].

Hence, cancer cells maintain telomere length because telomerase assures the addition of telomeric repeats to chromosomal ends. These cells have high telomerase activity whereas somatic cells do not [171, 172]. These findings are the base for the evaluation of telomerase inhibitors as potential anticancer agents [173-177]. The aim is to preserve from toxic side effects normal cells, which do not express the telomerase [178]. Different telomerase inhibitor strategies exist, based on targeting the different telomerase components or inhibiting telomerase activity.

```
                                                      hTR
                                                       ↓
5′GGTTAGGGTTAGGGGTTAG        3′                          — 5′
3′CCAAT        CAAUCCCAAUC                            hTERT
      |                              ↘ 3′
Addition of telomeric repeats
      ↓
5′GGTTAGGGTTAGGGTTAGGGTTAG
3′CCAAT        CAAUCCCAAUC                               — 5′
      |
Template repositioning                    ↘ 3′
      ↓
5′GGTTAGGGTTAGGGTTAGGGTTAG
3′CCAATCCC        CAAUCCCAAUC                            — 5′
      |
Addition of telomeric repeats              3′
      ↓
5′GGTTAGGGTTAGGGTTAGGGTTAGGGTTAG
3′CCAATCCC        CAAUCCCAAUC                            — 5′
                                          ↘ 3′
```
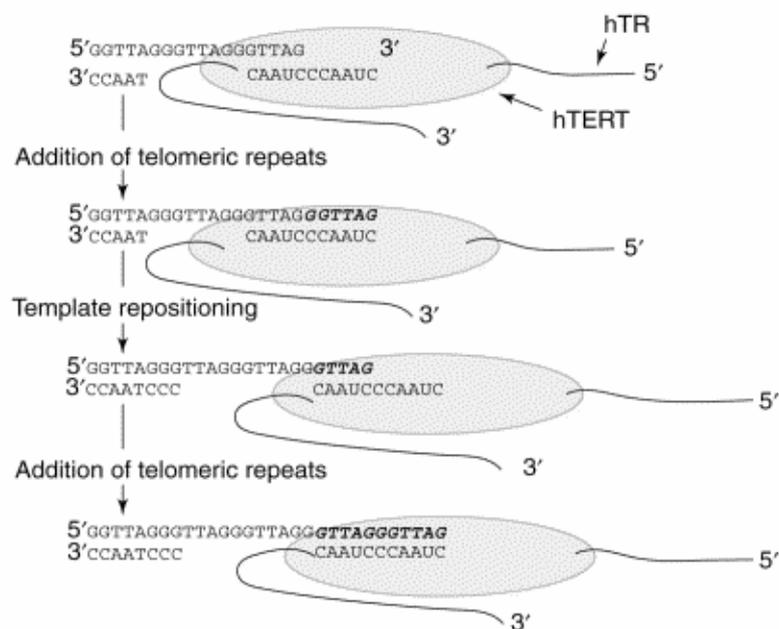
Fig. 3.2 - *Telomerase reaction cycle. Telomere binding involves base-pairing with the RNA template of hTR. In every reverse transcription cycle, the enzyme adds six bases, then translocates the template RNA for the synthesis of the next six bases.*

Several approaches to target telomeres and telomerase activity have been described [66, 154, 172, 176, 179-191]. One therapeutic strategy is based on immunotherapy which uses vaccines that support the immune system in their attack to cancer cells [192-194]; another therapeutic approach is based on targeting the hTR component of telomerase by antisense oligonucleotides [188, 195-197]; a further strategy known as "gene therapy" is based on the use of telomerase promoter-driven expression of suicide genes or oncolytic adenoviruses to selectively induce apoptosis in the telomerase-positive tumor cells [198-201]. A recent strategy is based on the use of hammerhead ribozymes, catalytic RNA molecules which cleave target RNAs in a site-specific manner, to target hTERT or hTR [202, 203].

Finally, an approach is addressed to design and synthesize drugs able to bind and stabilize the folding of the telomeric DNA G-rich overhangs into G-quadruplex structures. The formation of G-quadruplexes inhibits telomerase because the enzyme requires a single-stranded template [204]. Several ligands are able to promote the folding of the telomeric strand into G-quadruplexes. This folding inhibits telomerase function and induces cell senescence and/or apoptosis.

3.2 Telomeric quadruplex structures

The truncated sequence $d[AG_3(T_2AG_3)_3]$ from human telomeres has been first studied by NMR in Na^+ solutions [52] and by X-ray crystallographic studies in K^+ solutions [22]. The two G-quadruplex structures were found to be completely different (**Fig. 3.3**). In Na^+ solutions the sequence formed an intramolecular antiparallel structure with a diagonal TTA loop at one end and two lateral TTA loops at the other end (basket-type), whereas, the guanine glycosidic torsion angles in each tetrad were *syn-syn-anti-anti*. On the other hand, the crystal structure showed a parallel-stranded intramolecular quadruplex with three symmetrical propeller TTA loops and with the guanines glycosidic angles all in *anti* conformation. Later, sedimentation and fluorescence studies clearly showed that the structure of the human telomeric quadruplex obtained with crystallographic analyses was not the major conformation in K^+ solutions, as a plethora of conformations existed under these conditions [205]. Given that K^+ is more abundant than Na^+ in cells, many groups addressed their studies on topological arrangements of the intramolecular human telomeric G-quadruplex in K^+ solutions [205-210].
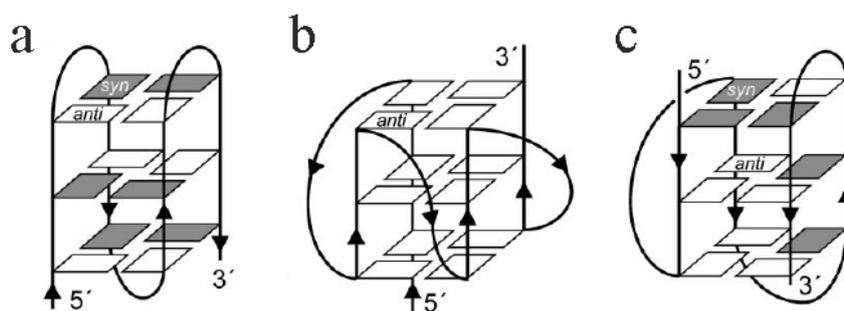
Fig. 3.3 - *Schematic representations of the folded structures of $d[AG_3(T_2AG_3)_3]$. (a) The Na^+ solution structure determined by NMR. (b) The K^+ crystal structure. (c) The folding topology in K^+ solution recently proposed by NMR and circular dichroism studies.*

In the 2006, four papers, appeared almost at the same time, have reported folding topologies of human telomeric G-quadruplex in K^+ solution [211-214]. Yang and co-workers reported an unprecedented hybrid-type G-quadruplex folding topology in K^+ solution, obtained by NMR, with mixed G-arrangements of G-tetrads that are connected with a double-chain-reversal side loop and two lateral loops (**Fig. 3.3**) [211]. The same structure was suggested by Sugiyama and co-workers on the base of CD spectra of the $d[AG_3(T_2AG_3)_3]$ sequence in which guanines were replaced with 8-bromoguanines. They also hypothesized that the hybrid-type structure found in K^+ solutions could favour the formation of t-loops at the end of chromosomes, providing the environmental conditions

for protein binding (**Fig. 3.4**) [214]. Patel and co-workers found by NMR the same unimolecular human telomeric structure in which G-tetrads present an *anti-syn-syn-syn* and two *syn-anti-anti-anti* guanine glycosidic torsion angles [212]. They described this G-quadruplex structure, in which three strands are oriented in one direction and the fourth is in the opposite direction, as a (3+1) topology. In another work, they have additionally shown that the human telomeric sequence can adopt two distinct intramolecular foldings containing the same (3+1) G-tetrad core, but different loop arrangements [213]. Recently, Yang and coworkers reported the NMR structure of the G-quadruplex formed by the 22-mer human telomeric sequence plus a flanking AA at each end [215]. This hybrid-type structure consists of three G-tetrads linked with mixed parallel-antiparallel G-strands and three TTA loops. The two bottom G-tetrads present the *anti-anti-syn-anti* arrangement of the guanine glycosidic torsion angles and the top G-tetrad presents the reversed *syn-syn-anti-syn* arrangement. Interestingly, they found a novel adenine triple platform formed by the three adenines of the wild-type sequence (A3, A9 and A21), capping the top tetrad of the quadruplex.
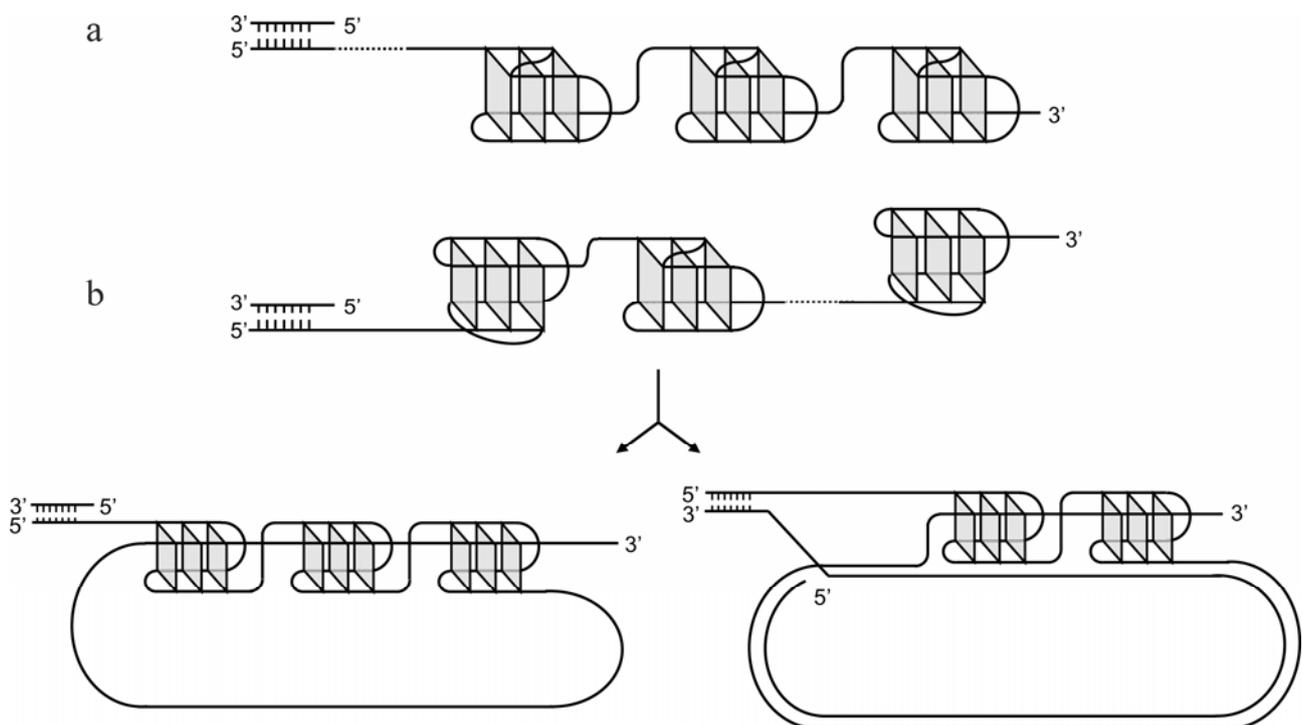
**Fig. 3.4** - *Models of multimers formed by the human telomeric DNA proposed by (a) Yang and co-workers and (b) Sugiyama and co-workers. They also proposed models showing t-loop formation with the mixed parallel/antiparallel quadruplex structures.*

The ability of telomeric G-quadruplexes to interconvert among different forms in different conditions could perhaps be the key of their biology. The presence of diverse G-quadruplex

topological arrangements could have important implications for the structure-based drug design and the understanding of DNA G-quadruplex structures and topologies is central to the design of ligands that can bind and stabilize them and thereby modulate particular cellular pathways and functions. Moreover, the formation and stability of these structures depend on the presence and on the type of monovalent cations; as a consequence, the presence of different cations could have an effect on the binding behaviour of drugs.

On these basis, the interactions of two investigated G-quadruplexes, specifically the unimolecular $d[AG_3(T_2AG_3)_3]$ quadruplex from the human telomeric sequence, and the tetramolecular quadruplex $[d(TGGGGT)]_4$ from the truncated telomeric sequence of *Oxytricha* and *Tetrahymena*, were examined using two different buffered solutions containing either K^+ or Na^+ at a fixed ionic strength, to establish any dependence for binding behaviour on the metal ion present in solution.

Fig. 3.5 shows the CD spectrum for the $d[AG_3(T_2AG_3)_3]$ quadruplex in Na^+ containing solution. The spectrum is characterized by a positive band at 296 nm and a negative band at 263 nm, that are characteristic of the antiparallel-stranded quadruplex structures [216, 217].
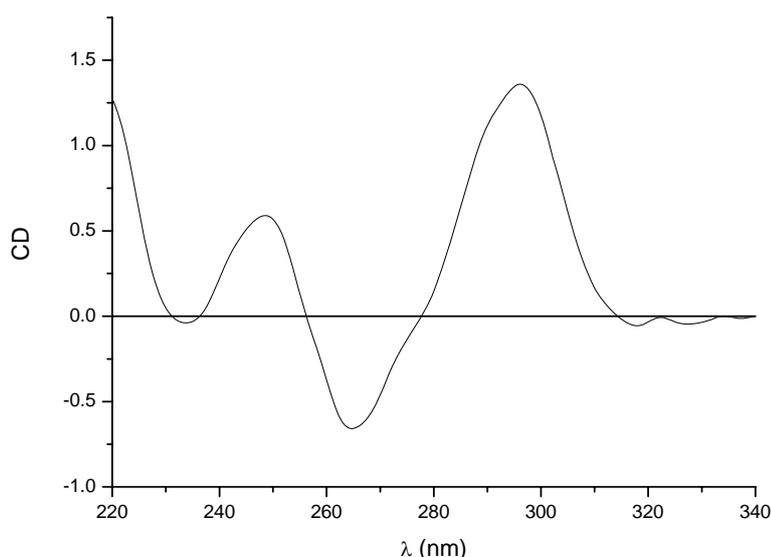
Fig. 3.5 – *CD spectrum for the d[AG₃(T₂AG₃)₃] quadruplex using the Na⁺ containing solution.*

The CD spectrum of the $d[AG_3(T_2AG_3)_3]$ quadruplex using K^+ solution is quite dissimilar to the spectrum of the same sequence in Na^+ buffer and it is characterized by two positive bands at 263 nm and 290 nm, respectively, and by a negative band at 238 nm (**Fig. 3.6**). This spectrum is characteristic of the (3+1) hybrid-type G-quadruplex folding topology.
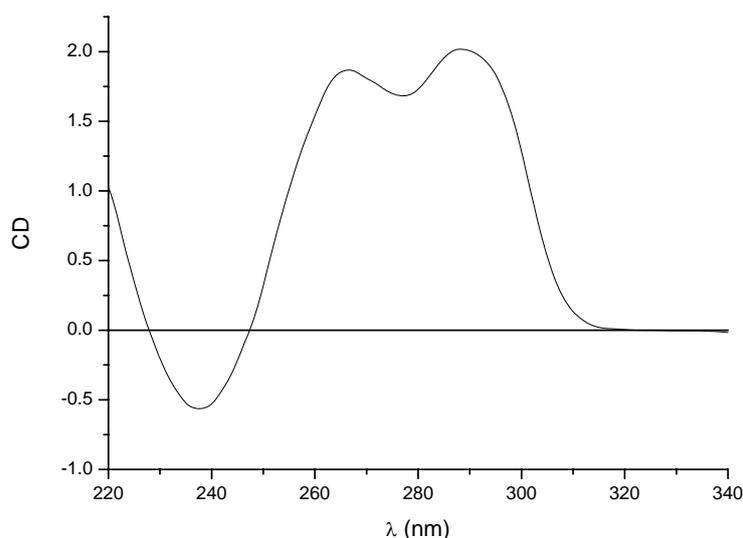
Fig. 3.6 – *CD spectrum for the d[AG₃(T₂AG₃)₃] quadruplex using the K⁺ containing solution.*

The truncated telomeric sequence from *Oxytricha* and *Tetrahymena*, d(TGGGGT), has been previously reported in literature to form, in presence of cations, a tetramer, with a parallel-stranded structure. This G-quadruplex structure has been solved by NMR and X-ray techniques, the strands associate generating a right-handed helix containing four equivalent grooves and all bases in the *anti* conformation [16, 36, 218].

Fig. 3.7 shows the CD specta of the [d(TGGGGT)]₄ quadruplex in the two different buffered solutions containing either K⁺ or Na⁺ cations. The spectra are quite similar, they are characterized by a intense positive band at 263 nm and a negative band at 243 nm that, as reported in literature, are characteristic of the parallel-stranded quadruplex structures [216, 217]. These qualitatively similar spectra confirm that this molecule really adopts a parallel-stranded structure in both experimental solution conditions.
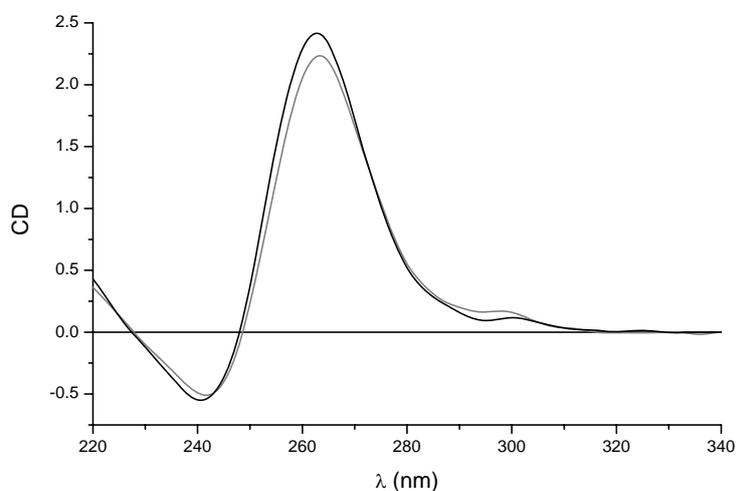
Fig. 3.7 - *CD spectra of the [d(TGGGGT)]₄ quadruplex in the K⁺ (black line) and Na⁺ (gray line) containing solution.*

3.3 Quadruplex-binding ligands

The quadruplex-binding ligands investigated in this thesis were the distamycin A (Dist-A); the compounds **1** and **2**, two derivatives of distamycin; and the cationic porphyrin TMPyP4.

Dist-A (**Fig. 3.8**) is a small molecule with antibiotic properties which binds with high affinity to the minor groove of B-form DNA [103] and has been shown to interact with four-stranded parallel DNA quadruplex containing oligonucleotides of different sequences [104, 105]. Two opposite models have been proposed for the distamycin-quadruplex complex: the first one suggests that distamycin molecules bind as dimers in two opposite grooves of quadruplex [d(TGGGGT)]$_4$ [104]; the second one suggests that two distamycin molecules stack on the terminal G-tetrad planes of the quadruplexes [d(TAGGGTTA)]$_4$, [d(TAGGGGTT)]$_4$ and [d(TAGGGGGT)]$_4$ [105].

Compounds **1** and **2** are two carbamoyl analogues of distamycin A, containing four and five pyrrole units, respectively (**Fig. 3.8**).

Derivatives of Dist-A have been reported to be inhibitors of the human telomerase enzyme [219]. More recently, some distamycin analogues have been found interacting with a DNA quadruplex of sequence d[GGG(TTAGGG)$_3$]; a mixed groove/G-quartet stacking binding mode was suggested for the interaction on the basis of molecular modelling [220].
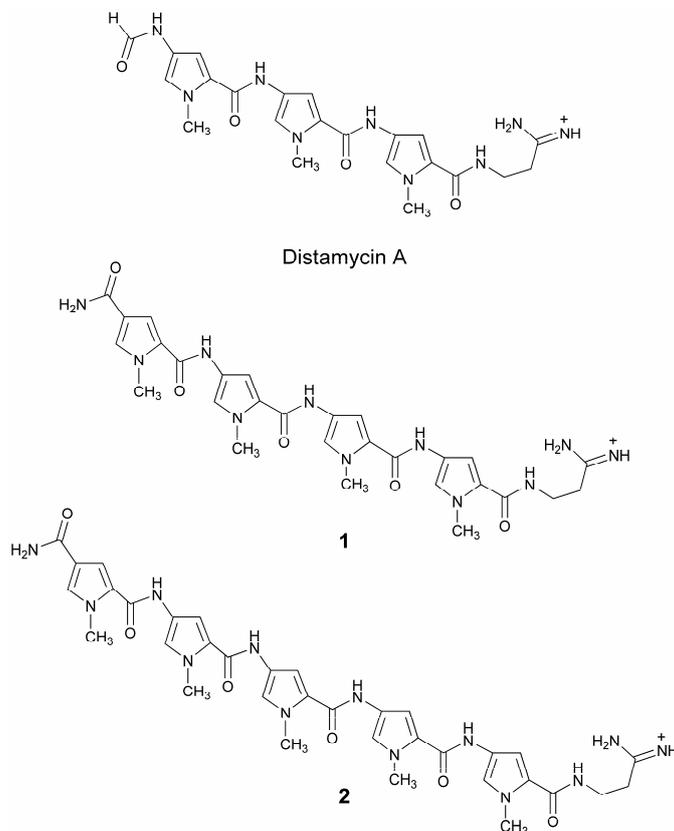
Fig. 3.7 - *Chemical structures of Dist-A and its derivatives.*

The cationic porphyrin TMPyP4 (**Fig. 3.8**) has been extensively studied as a quadruplex-binding ligand since it induces telomerase inhibition upon binding to telomeric DNA quadruplexes [221, 222]. TMPyP4 was supposed to be of appropriate size to stack with the G-tetrads that stabilize quadruplex DNA, however, it does not have high selectivity for DNA quadruplexes compared to duplex DNA [223, 224].

The stoichiometry and nature of the structure of porphyrin-quadruplex complexes have been controversial, and various models have been proposed, in large part on the basis of spectroscopic studies. These models involve TMPyP4 molecules in intercalative binding between adjacent G-tetrads [225], at a G-A interface [226], or stacked externally onto the end of a quadruplex [225, 227].
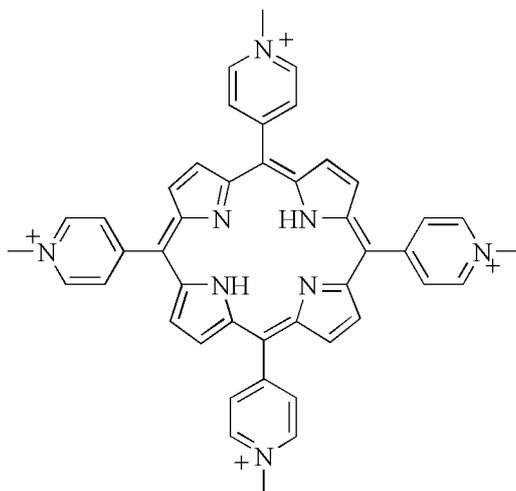
Fig. 3.8 - *Chemical structure of TMPyP4.*

Recently, Parkinson et al. reported the crystal structure of the bimolecular human telomeric quadruplex [d(TAGGGTTAGGG)]₂, in complex with TMPyP4 molecules (**Fig. 3.9**) [228]. The porphyrin molecules bind by stacking onto the TTA nucleotides, either as part of the external loop structure or at the 5' region of the stacked quadruplex. This structure defines a novel binding mode for the porphyrin TMPyP4, since there are no direct ligand interactions with G-tetrads. This is in accord with the relative nonselectivity by TMPyP4 for quadruplexes compared to duplex DNA.
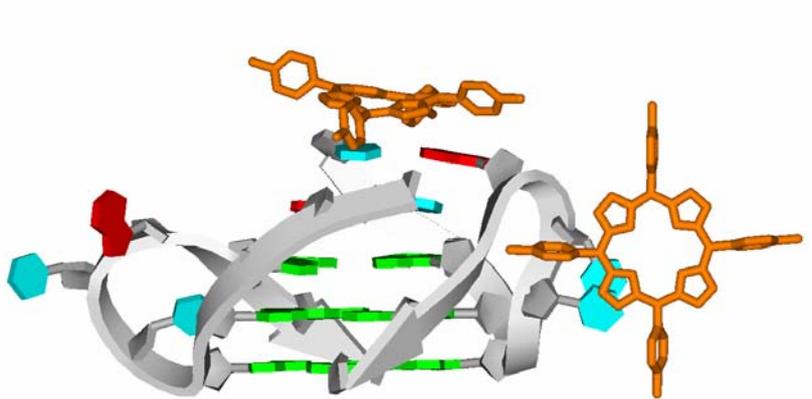
Fig. 3.9 - *Crystal structure of the [d(TAGGGTTAGGG)]₂ quadruplex in complex with TMPyP4 molecules*

3.4 Materials and methods

The oligonucleotides used were purchased from Sigma-Aldrich. DNA solutions were prepared by dissolving solid lyophilized oligonucleotides in buffered solutions containing either K^+ or Na^+. The K^+ buffer used was 10 mM potassium phosphate, 70 mM KCl, 0.1 mM EDTA; the Na^+ buffer was 10 mM sodium phosphate, 70 mM NaCl, 0.1 mM EDTA; both buffer solutions were at pH= 7.0. DNA quadruplexes were formed by heating the solutions to 90 °C for 5 min. The solutions were then cooled slowly to room temperature and equilibrated for one day at 4 °C. The oligonucleotides concentrations were determined by UV adsorption measurements at 90°C using molar extinction coefficient values $\varepsilon_{(260\ nm)}$ of 57800, 116300, 234500 $M^{-1}cm^{-1}$ for d(TGGGGT), d(TAGGGTTAGGG), and d[$AG_3(T_2AG_3)_3$], respectively. The molar extinction coefficients were calculated by the nearest neighbor model [229].

TMPyP4 porphyrin and distamycin A were purchased from Sigma-Aldrich; compounds **1** and **2** were synthesized as previously described [230]. Ligand solutions were freshly prepared every time before binding experiments by dissolving the drugs in the same K^+ or Na^+ buffer used for DNA quadruplexes.

3.4.1 Isothermal titration calorimetry

ITC experiments were performed on a CSC 4200 Calorimeter from Calorimetry Science Corporation (CSC, Utah) with a cell volume of 1.3 mL. Calorimetric titrations were carried out at 25 °C, by injecting 15 µL aliquots of a 300-940 µM ligand solution into a 30-50 µM quadruplex solution at 400 seconds intervals, with stirring at 297 rpm, for a total of 16 injections.

Control experiments were performed by injecting each ligand in the buffers to obtain the heat effects associated to dilution of each drug. The calorimetric enthalpy for injection was calculated, after correction for the dilution heat of drug, by integration of the areas under the injection peaks. To obtain the thermodynamic parameters of the interactions, the corrected heat values were plotted against the molar ratio and analyzed using the Bindwork program supplied with the instrument according to the one- and two-binding site models.

3.5 Results of ITC study

We used ITC to characterize quantitatively the thermodynamics of drugs binding to quadruplexes. ITC measures heat generated or absorbed upon binding, and provides in a single experiment the values of the binding constant (K_b), the stoichiometry (n) and the change of enthalpy ($\Delta H°$). The K_b value then permits calculation of the change in free energy ($\Delta G°$), which together with $\Delta H°$ allows the calculation of the entropic term T·$\Delta S°$. The separation of the free energy of binding into enthalpic and entropic components allows to reveal the nature of the forces that drive the binding reaction. The interactions of distamycin and its derivatives with the [d(TGGGGT)]$_4$ and d[AG$_3$(T$_2$AG$_3$)$_3$] quadruplexes were examined at 25 °C using two different buffered solutions containing either K$^+$ or Na$^+$ at a fixed ionic strength. Since thermodynamic parameters associated to the interaction of TMPyP4 with [d(TGGGGT)]$_4$ and d[AG$_3$(T$_2$AG$_3$)$_3$] quadruplexes were already reported in literature, in this thesis was examined only the binding of TMPyP4 to the bimolecular quadruplex [d(TAGGGTTAGGG)]$_2$ using the K$^+$ containing solution.

Fig. 3.10 and **Fig. 3.11** show the results of calorimetric titrations concerning the interactions of [d(TGGGGT)]$_4$ with distamycin and compound **1**, respectively. **Fig. 3.12** shows the results for the titration of d[AG$_3$(T$_2$AG$_3$)$_3$], and **Fig. 3.13** shows the results of calorimetric titration of TMPyP4 with [d(TAGGGTTAGGG)]$_2$.

Upper panels of the figures show typical ITC profiles for the binding of ligands to the quadruplexes. Each of the heat burst curves in the figures corresponds to a single drug injection. The areas under these heat burst curves were determined by integration to yield the associated injection heats. The interaction heats were corrected for the heat of dilution associated with each ligand, independently determinated by injecting ligand solution into the buffers. The heat of dilution for the distamycin was found to be slightly endothermic, while the dilution of TMPyP4 and of the derivatives of distamycin was moderately exothermic. The corrected heat values are plotted as a function of the molar ratio in the lower panels of the figures, to give the corresponding binding isotherms. The resulting isotherms were then fitted using the one- or two-binding site model to give the binding enthalpy ($\Delta H°$), equilibrium binding constant (K_b), and stoichiometry (n). The remaining thermodynamic parameters, $\Delta G°$ and T$\Delta S°$, were derived using the standard relationships $\Delta G° = -RT \ln K_b$ and T$\Delta S° = \Delta H° - \Delta G°$. Resolvable binding isotherms were never obtained for the compound **2** in both solution conditions using any combination of reactant concentration, suggesting a poor affinity of the molecule for the investigated quadruplexes. The thermodynamic parameters determined for the binding to [d(TGGGGT)]$_4$ and d[AG$_3$(T$_2$AG$_3$)$_3$] are reported in **Table I** and **II**, respectively. The thermodynamic parameters determined for the interaction of

TMPyP4 are collected in **Table III**, along with the already reported parameters for an easy comparison.

3.5.1 ITC data for [d(TGGGGT)]$_4$

Examination of the thermodynamic data reveals that distamycin–[d(TGGGGT)]$_4$ interactions are exothermic. Interaction of Dist-A in the K$^+$ buffer gave rise to a more complex binding isotherm (**Fig. 3.10**) with a curvature typically observed in the presence of multiple, non-equivalent binding sites. The biphasic nature of the isotherm suggested two binding sites with different affinities. Indeed, the ITC data were best fit using a two-site model whose thermodynamic parameters are listed in **Table III**. The binding stoichiometry is 2:1 and 4:1 (drug:quadruplex) for the first and second binding event, respectively. The affinity between Dist-A and [d(TGGGGT)]$_4$ in the K$^+$ buffer is enhanced when the ratio of Dist-A to the quadruplex is increased, as the equilibrium constant for the second interaction event is ~10-fold greater than that for the first event. Particularly, K_b values of 4 x 10^5 M^{-1} and 4 x 10^6 M^{-1} have been found for 2:1 and 4:1 complexes, respectively.
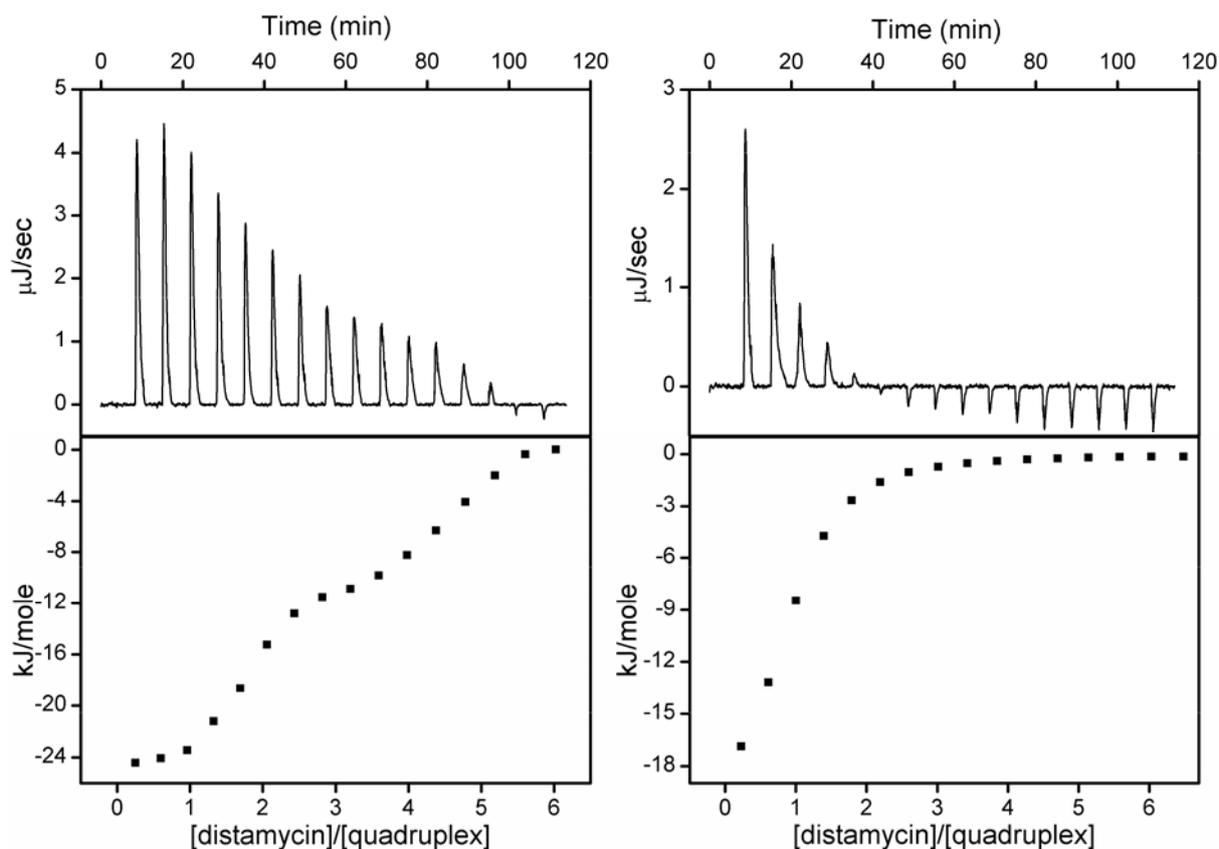
Fig. 3.10 - *ITC data for the binding of distamycin to [d(TGGGGT)]$_4$. Top panels show the calorimetric response for injections of distamycin to [d(TGGGGT)]$_4$ in the K$^+$ containing solution (left) and Na$^+$ containing solution (right), pH 7.0 and 25 °C. Bottom panels show the integrated injections heats for the above data.*

44

The binding behaviour of Dist-A in the Na^+ containing solution was found to be considerably different compared to the equivalent K^+ containing solution (**Fig. 3.10**). In this case there is only one binding event with a stoichiometry of 1:1 and an association constant of 2×10^5 M^{-1} that is lower than the constants observed in K^+ buffer.

The raw data for the titrations of compound **1** with $[d(TGGGGT)]_4$ (**Fig. 3.11**, upper panels) indicate an endothermic interaction, based on the negative values observed for the peaks. With each injection of ligand, less and less heat uptake was observed until constant values were obtained (corresponding to the heat of dilution), reflecting a saturable process.
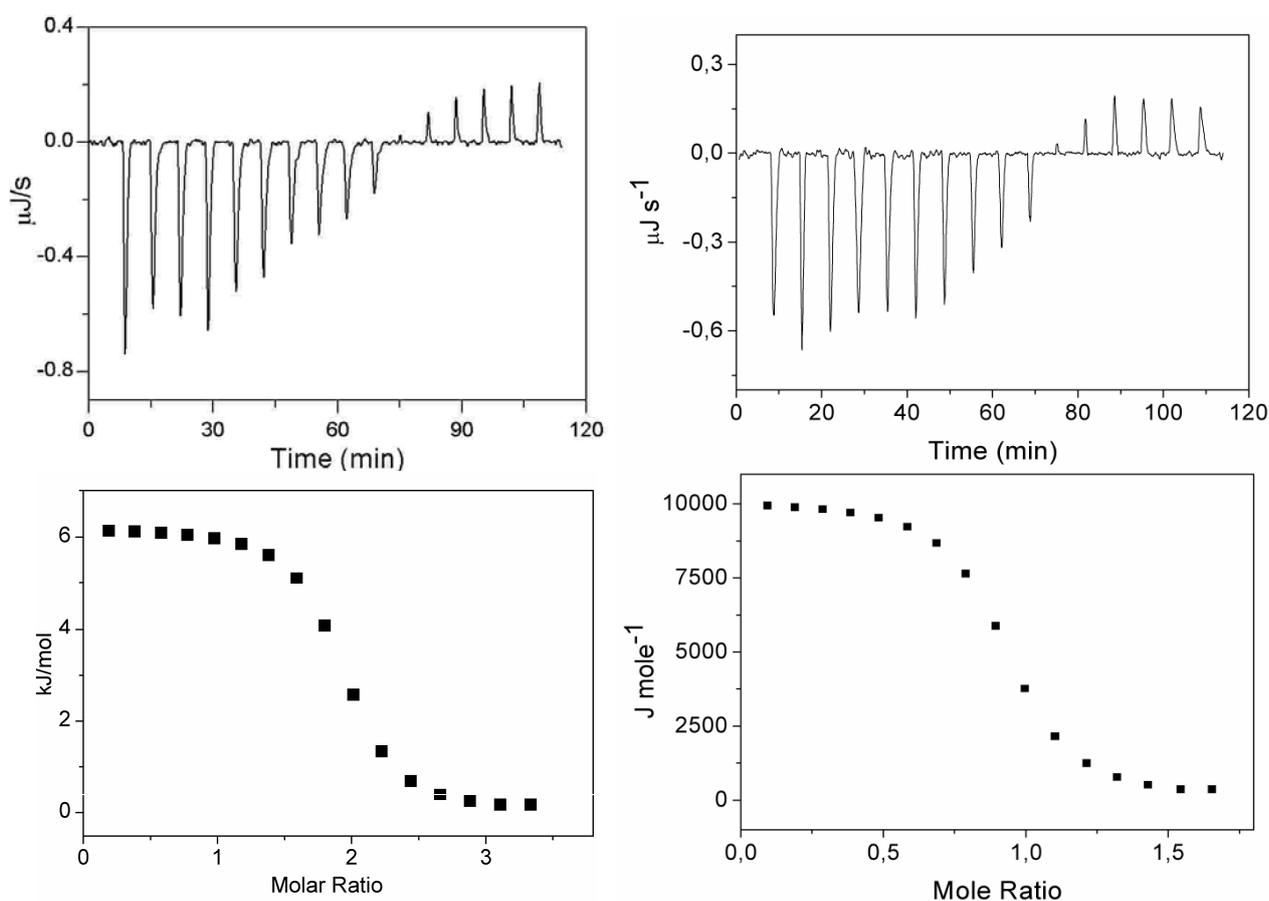
Fig. 3.11 - *ITC data for the binding of **1** to $[d(TGGGGT)]_4$. Top panels show the calorimetric response for injections of **1** to $[d(TGGGGT)]_4$ in the K^+ containing solution (left) and Na^+ containing solution (right), pH 7.0 and 25 °C. Bottom panels show the integrated injections heats for the above data.*

The analysis of the thermodynamic data (**Table I**) shows that the **1**-$[d(TGGGGT)]_4$ interaction in the K^+ buffer is characterized by a $K_b = 2 \times 10^6$ M^{-1}, with a binding stoichiometry of 2:1 (drug:quadruplex). Moreover, the binding of **1** to the $[d(TGGGGT)]_4$ quadruplex was characterized by a moderately large and positive enthalpy of binding (7.0 kJ mol^{-1}). Except for stoichiometry,

similar results were obtained for the interaction of compound **1** with [d(TGGGGT)]$_4$ in Na$^+$ solution (**Table I**). The positive heat of formation of the **1**-[d(TGGGGT)]$_4$ complex indicates that this interaction is an enthalpically unfavourable (entropically driven) process in both solution conditions.

Table I
Thermodynamic parameters for the interaction of Dist-A and **1** with [d(TGGGGT)]$_4$ determined by ITC at 25 °C and pH 7.0

Buffer	Ligand	n	K_b (10^6 x M^{-1})	$\Delta_b H°$ (kJ/mol)	$T\Delta_b S°$ (kJ/mol)	$\Delta_b G°$ (kJ/mol)
K$^+$	Dist-A [a]	1.8 ± 0.2	0.4 ± 0.3	-8.0 ± 1.0	24 ± 2	-32 ± 2
		4.2 ± 0.2	4.0 ± 3.0	-10.0 ± 1.0	27 ± 2	-37 ± 2
	1	1.9 ± 0.1	2.0 ± 1.0	7.0 ± 0.5	43 ± 1	-36 ± 1
Na$^+$	Dist-A	1.0 ± 0.1	0.2 ± 0.1	-14.0 ± 0.5	16 ± 1	-30 ± 1
	1	0.9 ± 0.1	2.3 ± 1.0	10.0 ± 0.5	46 ± 1	-36 ± 1

[a] two binding events

3.5.2 ITC data for d[AG$_3$(T$_2$AG$_3$)$_3$]

The values of the binding constants and the Gibbs energy changes from ITC measurements (**Table II**) indicate that the associations between the distamycin and compound **1** with the d[AG$_3$(T$_2$AG$_3$)$_3$] quadruplex are favoured, at 25 °C, from thermodynamic point of view. Although the binding stoichiometry is 1:1 for all interactions (**Fig. 3.12**), the binding behaviours in the Na$^+$ solution were found to be considerably different compared to the equivalent K$^+$ solutions, with a different ligand affinity and/or enthalpy of binding. Indeed, although the affinity of **1** for the quadruplex is, in both solution condition, higher than distamycin, the equilibrium constants for the interactions of distamycin and **1** with d[AG$_3$(T$_2$AG$_3$)$_3$] in Na$^+$ buffer (K$_b$ = 1 x 10^6 M^{-1} and 6 x 10^6 M^{-1}, respectively) are higher than the corresponding constants in the K$^+$ buffer (K$_b$ = 6 x 10^5 M^{-1} and 2 x 10^6 M^{-1}, respectively). The values of ΔH° and ΔS° show that, in all cases, the binding processes are entropically driven; however, the interaction of distamycin with d[AG$_3$(T$_2$AG$_3$)$_3$] is always associated with a favourable binding enthalpy (-9 and -14 kJ mol^{-1} for Na$^+$ and K$^+$ solution, respectively), whereas the binding of **1** is exothermic (ΔH° = -18 kJ mol^{-1}) in the Na$^+$ buffer and moderately endothermic (ΔH° = 9 kJ mol^{-1}) in the K$^+$ solution.

3.5.3 ITC data for [d(TAGGGTTAGGG)]$_2$

Typical ITC results on the TMPyP4-[d(TAGGGTTAGGG)]$_2$ interaction at 25 °C along with the binding isotherm generated from the ITC data are shown in **Fig. 3.13**. Calorimetric data were fitted to a single set of identical sites model, and the thermodynamic parameters obtained from the best fit

are listed in **Table III**. The reaction which is exothermic, as also visible from the ITC profile, proceeds by little change in entropy (T·ΔS° = -5 kJ mol⁻¹) and yields a binding site value of 2. The experimental value of ΔH° is -36 kJ mol⁻¹ and K_b is 3 x 10⁵ M⁻¹. Although the binding stoichiometry is the same found for the already reported interactions of TMPyP4 with [d(TGGGGT)]₄ [231] and d[AG₃(T₂AG₃)₃] [232], the value of ΔH° is higher than the enthalpy changes previously observed (**Table III**).
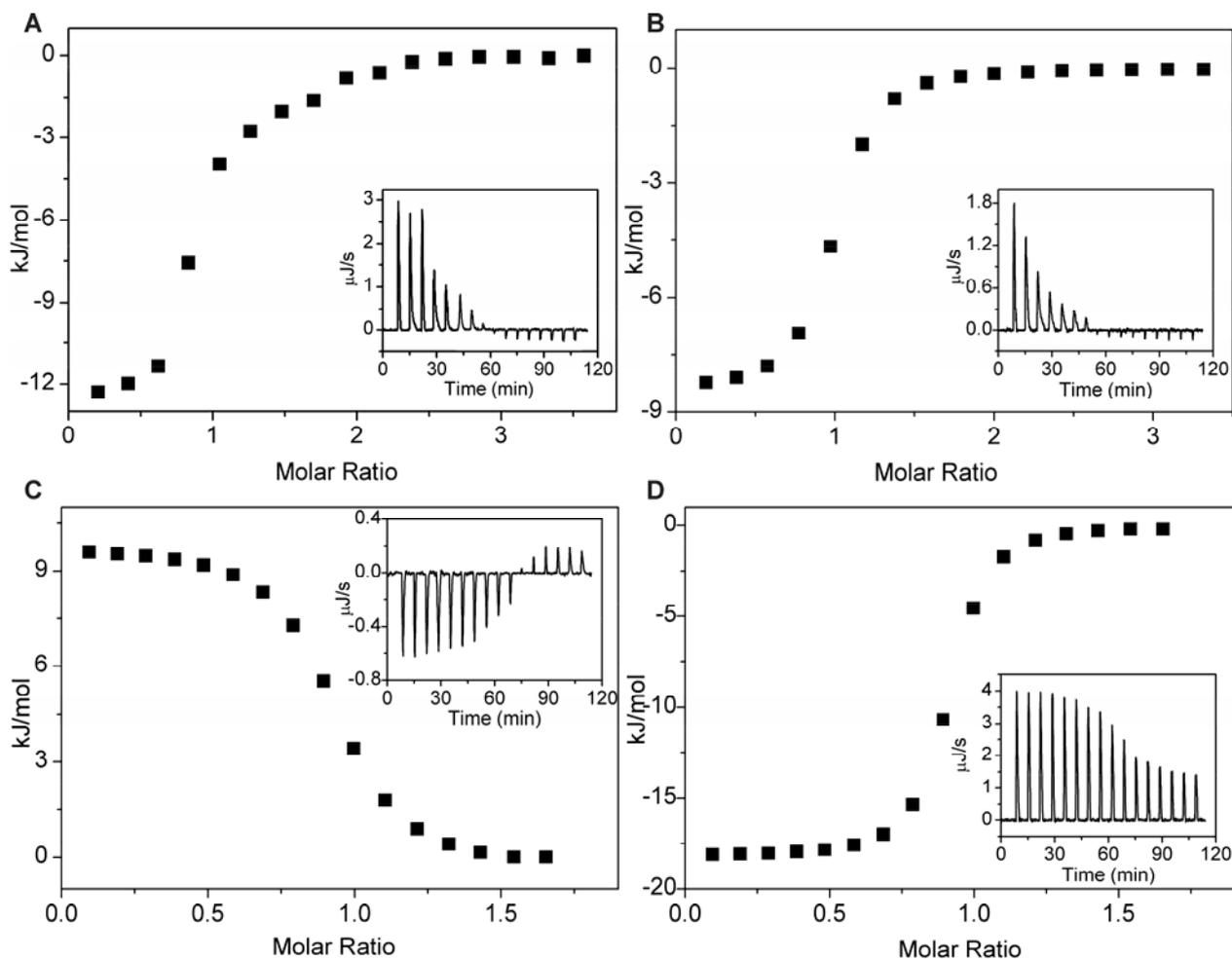
Fig. 3.12 - *ITC data curves (inserts) and binding isotherms for titration of d[AG₃(T₂AG₃)₃] with Dist-A in K⁺ (A) and Na⁺ (B) buffer and with compound **1** in K⁺ (C) and Na⁺ (D) buffer.*

Table II

Thermodynamic parameters for the interaction of Dist-A and **1** with d[AG₃(T₂AG₃)₃] determined by ITC at 25 °C and pH 7.0

Buffer	Ligand	n	K_b (10⁶ x M⁻¹)	$\Delta_b H°$ (kJ/mol)	$T\Delta_b S°$ (kJ/mol)	$\Delta_b G°$ (kJ/mol)
K⁺	Dist-A	0.9 ± 0.1	0.6 ± 0.3	-14.0 ± 2.0	19 ± 2	-33 ± 1
	1	1.0 ± 0.1	2.0 ± 1.0	9.0 ± 1.0	45 ± 1	-36 ± 1
Na⁺	Dist-A	1.0 ± 0.1	1.0 ± 0.5	-9.0 ± 1.0	25 ± 1	-34 ± 1
	1	0.9 ± 0.1	6.0 ± 3.0	-18.0 ± 1.0	21 ± 1	-39 ± 1
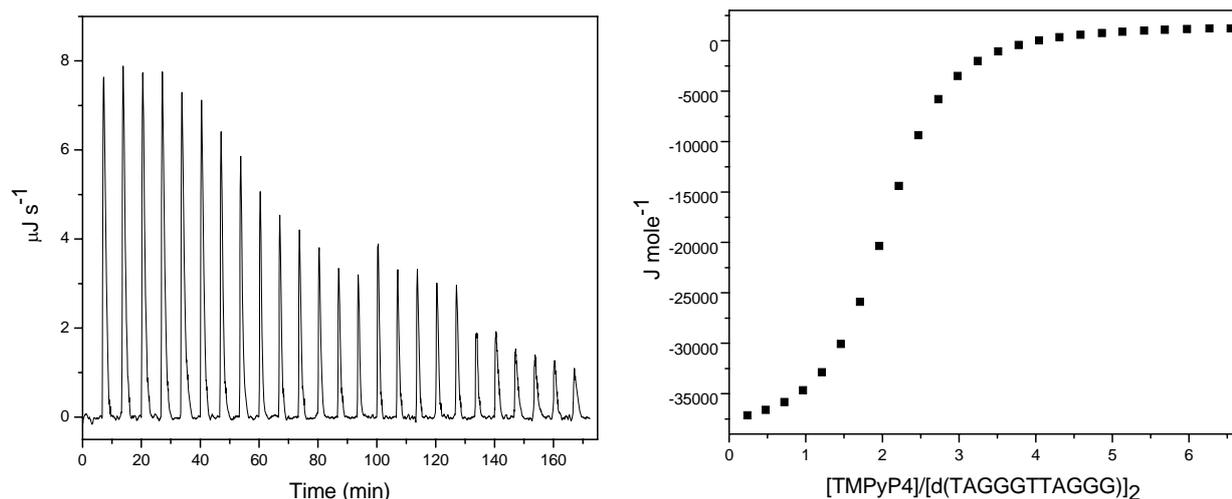
Fig. 3.13 - *ITC data for the binding of TMPyP4 to [d(TAGGGTTAGGG)]₂. Left panel shows the calorimetric response for injections of TMPyP4 to a [d(TAGGGTTAGGG)]₂ solution at pH 7.0 and 25 °C. Right panel shows the integrated injections heats for the data.*

Table III

Thermodynamic parameters for the interaction of TMPyP4 in K^+ buffer determined by ITC at 25 °C and pH 7.0

Quadruplex	n	K_b (10^5 x M^{-1})	$\Delta_b H°$ (kJ/mol)	$T\Delta_b S°$ (kJ/mol)	$\Delta_b G°$ (kJ/mol)
[d(TAGGGTTAGGG)]₂	2.0 ± 0.1	3 ± 1	-36 ± 2	-5 ± 1	-31 ± 1
[d(TGGGGT)]₄ [a]	2.0 ± 0.1	2.5 ± 0.2	-23.9 ± 2.8	6.9 ± 2.8	-30.8 ± 0.2
d[AG₃(T₂AG₃)₃] [b]	1.9 ± 0.4	0.28 ± 0.07	-17.6 ± 3.3	8.0 ± 3.3	-25.5 ± 0.4

[a] from ref. [233]
[b] from ref. [234]

3.6 Discussion

Numerous structural and thermodynamic studies have been carried out in recent years in order to characterize the binding of quadruplex-interactive compounds [228, 232-235]. Besides those, in this thesis is reported the calorimetric study of the interaction of distamycin A and its derivatives **1** and **2** with the [d(TGGGGT)]$_4$ and d[AG$_3$(T$_2$AG$_3$)$_3$] quadruplexes and of TMPyP4 with [d(TAGGGTTAGGG)]$_2$. Experiments reveal that TMPyP4 binds [d(TAGGGTTAGGG)]$_2$, as well as distamycin A and compound **1** bind the [d(TGGGGT)]$_4$ and d[AG$_3$(T$_2$AG$_3$)$_3$] quadruplexes, conversely the compound **2** appears to have a very poor affinity in any case.

The thermodynamic parameters determined by ITC (**Tables I** and **II**) indicate that the interactions of both Dist-A and **1** with quadruplexes are entropically driven processes in both solution conditions. However, calorimetric experiments reveal a notable alteration in the binding behaviour of the drugs upon switching from the K$^+$ to Na$^+$ buffer.

The interaction of distamycin with [d(TGGGGT)]$_4$ in the K$^+$ buffer give rise to a biphasic binding isotherm with a curvature typically observed in the presence of two binding sites with different affinities [236]. Indeed, the ITC data were best fit using a two-site model whose thermodynamic parameters are listed in **Table I**. The binding stoichiometry is 2:1 and 4:1 (drug/quadruplex) for the first and second binding event, respectively. Interestingly, the affinity between Dist-A and [d(TGGGGT)]$_4$ in the K$^+$ buffer is enhanced when the ratio of distamycin to the quadruplex is increased, as the equilibrium constant for the second interaction event is ~10-fold greater than that for the first event. It is interesting to note that interaction between Dist-A and the [d(TGGGGT)]$_4$ is entropically driven, being the -TΔS° value higher than ΔH° value. These data confirm once again the groove-binder nature of Dist-A, in fact, a recent analysis made on a number of DNA binding agents revealed that groove-binding is entropically driven [237].

These calorimetric results on the interaction of Dist-A with [d(TGGGGT)]$_4$ in K$^+$ buffer are confirmed by the three-dimensional structure of the complex solved by NMR. The complex indicate that distamycin molecules interact with [d(TGGGGT)]$_4$ in a 4:1 binding mode, with two antiparallel distamycin dimers that bind two opposite grooves of the quadruplex (**Fig. 3.14**) [236].

The binding behaviour of distamycin in the Na$^+$ containing solution was found to be surprisingly different compared to the equivalent K$^+$ containing solution. In this case, ITC experiments show that there is only one binding event with a stoichiometry of 1:1 and that the affinity of Dist-A is lower than the ones observed in K$^+$ buffer.

In the case of the interaction of compound **1** with [d(TGGGGT)]$_4$, similar results were obtained in both solution conditions, except for stoichiometry. Indeed, the calorimetric results obtained in K$^+$

solution reveal a stoichiometry of 2:1 (drug/quadruplex), whereas in the presence of Na^+ the stoichiometry is 1:1.
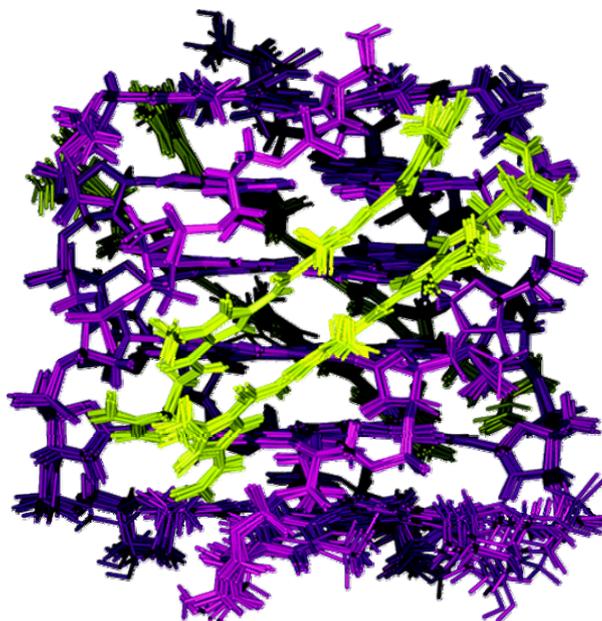
Fig. 3.14 - *Side view of the superimposition of the 10 best structures of the 4:1 complex Dist-A/[d(TGGGGT)]₄. Dist-A is reported in yellow, and DNA is colored in magenta.*

Interestingly, as shown in **Table I**, the association processes of both ligands with [d(TGGGGT)]₄ are always entropically driven, even if the direct ITC measurements of the binding enthalpies indicate that the enthalpic contribution for the binding of Dist-A favors the interaction, whereas for compound **1** the major entropic binding forces are countered by a small unfavorable enthalpic term.

In general, the binding of an intercalator to DNA is entirely driven by a large favorable enthalpy change but with an unfavorable entropy decrease, and the binding of a groove binder to DNA is driven by a large favorable increase in entropy [238, 239]. Clearly, it is not possible to distinguish possible binding modes of compound **1** from the available calorimetric titration data alone, but the similarity of thermodynamic data obtained for the distamycin (entropically driven processes) provides compelling support for a groove-binding mechanism also in this case.

Both of these drugs show a different behaviour in the binding of [d(TGGGGT)]₄ when the interaction occurs in the presence of K^+ or Na^+. Apparently, there is no specific reason to justify this disparity because the same parallel-stranded conformation is found in solution for the [d(TGGGGT)]₄ independently from the ions present [17, 36, 218, 240, 241].

However, a difference between the K^+ and Na^+ structures of [d(TGGGGT)]₄ involves the coordination of the cations. Indeed, typically three K^+ ions bind the quadruplex, placed exactly at the midpoint between two G-tetrad layers, whereas four Na^+ ions usually bind the [d(TGGGGT)]₄,

placed closer to the plane of the G-tetrad rather than between two G-tetrads [17, 32]. Additionally, the smaller Na^+ ions are quite mobile within the central cavity of the quadruplex and can even enter or exit from the quadruplex core. On the other hand, the larger K^+ ions remain caged in the central position between G-tetrads [242]. Interestingly, the reported NMR structure of the dystamycin-[d(TGGGGT)]$_4$ complex (**Fig. 3.14**) shows that the two dystamycin dimers are not perfectly centered into the quadruplex grooves, but they are slightly shifted towards the 5'-edge [236], and, reasonably, the different cations coordination could alter the recognition process of Dist-A. In fact, the presence of cations coordinated at the top and at the bottom of the cavity at the carbonyl oxygens of the G-tetrads, or their absence, could also influence the organization of the terminal thymines that (i) can occupy an external disordered position or (ii) form interstrand T-T base pairs or (iii) can even penetrate into the grooves forming van der Waals interactions and/or hydrogen bonds, as in some cases found in presence of Na^+ and Tl^+, preventing the binding of the ligands [240].

These effects on the binding may also partly reflect inherent differences in hydration for the quadruplex structures and/or binding-induced release of counterions. Indeed, the different ionic radii between the two cations could led to a different extent of ions' distribution all around the surface of the quadruplex [243]. In particular, the Na^+ ions should fit better into the grooves of the quadruplex, preventing or at least altering the binding of the groove ligands.

The calorimeric data for the interactions of distamycin and compound **1** with the d[AG$_3$(T$_2$AG$_3$)$_3$] reveal that, also in this case, the binding behaviour of the two drugs in the Na^+ solution is different compared to the equivalent K^+ solution. Although the overall stoichiometries are unaffected, different ligand affinity and/or enthalpy of binding were found. In particular, both ligands show a higher affinity for the quadruplex in the Na^+ buffer. Additionally, the binding of **1** in this buffer was characterized by a moderately large and positive enthalpy of binding, in contrast with the behaviour observed in the K^+ buffer.

This different behaviour in the two buffers could be more easily justified because the d[AG$_3$(T$_2$AG$_3$)$_3$] forms two completely different structures in Na^+ or K^+ solution [52, 212]. Indeed, it is well known that the human telomeric sequence forms an anti-parallel arrangement in Na^+ solution, with one diagonal and two lateral loops, whereas, in K^+ solution, this sequence appears to be able to adopts two distinct intramolecular foldings containing the same (3+1) G-tetrad core, but different loop arrangements both with one strand-reversal and two lateral loops [213].

The major entropic contribution for the binding of the ligands to the human telomeric quadruplexes suggests a groove-binding mechanism also for these interactions. However, the human telomeric quadruplexes used in this study differ from [d(TGGGGT)]$_4$ for the structure (grooves size and

presence of lateral loops), and for the number of stacked G-tetrad planes ([d(TGGGGT)]$_4$ and d[AG$_3$(T$_2$AG$_3$)$_3$] contain four and three planes, respectively). This could give explanation for the different stoichiometries in the binding of the [d(TGGGGT)]$_4$ and d[AG$_3$(T$_2$AG$_3$)$_3$] quadruplexes. Very interestingly, this ITC study shows that the structural modifications of the compounds **1** and **2** influence the energetic of interaction with quadruplexes, increasing (for compound **1**) or decreasing (for compound **2**) the affinity of the ligands toward the quadruplex molecules.

CHAPTER 4: STABILITY AND BINDING PROPERTIES OF A MODIFIED THROMBIN BINDING APTAMER

4.1 Thrombin binding aptamers

α-Thrombin is a serine protease with multiple functions in homeostasis, and it is the only protein capable to catalyze the cleavage of fibrinogen to produce fibrin clot [244]. The multiple functions of α-thrombin in hemostasis concern the procoagulant, anticoagulant and fibrinolytic pathways and involve a large number of substrates [245]. The crystallographic structure of α-thrombin showed structural features explaining the multiple functions of thrombin [246]. Two positively charged regions are present on the enzyme surface, termed exosites I and II, on the opposite sides with respect to catalytic site. Exosite I is usually referred to as the fibrinogen recognition site and exosite II as the heparin-binding site.

The malfunctioning of thrombin results in hemorrhage and an excessive coagulation function results in dissemination of the clot in not-damaged tissues, causing thrombosis [247]. The possibility to specifically inhibit thrombin *in vivo* with synthetic compounds is an important goal to prevent thrombosis.

In 1992, Bock *et al*. screened a pool of about 10^{13} synthetic oligonucleotides [248] and discovered a potent inhibitor of thrombin [99, 100] based on a single-stranded 15-mer DNA with the sequence $^{5'}$GGTTGGTGTGGTTGG$^{3'}$ (Thrombin Binding Aptamer, TBA).

The three-dimensional structure of TBA was solved by NMR spectroscopy [249, 250]. TBA forms a unimolecular quadruplex in solution, arranged in a chair-like structure, consisting of two G-quartets connected by two TT loops and a single TGT loop (**Fig. 4.1**). TBA does not interact with the active site of thrombin and a binding site was localized to the anion exosite I [251-254]. However, crystallographic studies of TBA complex with thrombin revealed that TBA may interact with both exosites I and II of thrombin [255, 256]. The three-dimensional structures of thrombin and TBA are preserved in the complex and specific interactions are found involving the exosites and the loops of the aptamer [256-258].

Recently, the TBA–based oligodeoxynucleotide $^{3'}$GGT$^{5'}$-$^{5'}$TGGTGTGGTTGG$^{3'}$ (mTBA), containing a 5'-5' site of polarity inversion, was synthesized with the aim to improve the biological and biophysical properties of the thrombin aptamer. The mTBA was studied by CD and NMR spectroscopy and its three-dimensional structure was obtained [106]. The structure presents a chair-like conformation characterized by an unusual folding with three strands parallel and one strand oriented in antiparallel manner (**Fig. 4.1**).

In this thesis, a differential scanning calorimetric (DSC) study of thermodynamic stability of TBA and mTBA is reported, along with an isothermal titration calorimetric (ITC) study to assess the binding stoichiometry and to gain information on the energetics of binding of TBA and modified TBA to the thrombin. In order to explain the different behavior of the two aptamers we performed Molecular Dynamics (MD) simulations of TBA and mTBA. In addition, to improve the understanding of the aptamer-thrombin recognition, a model of the complexes has been generated by means of docking calculations and subjected to MD simulations in explicit water. Calorimetric data are discussed also on the basis of the MD simulations results.
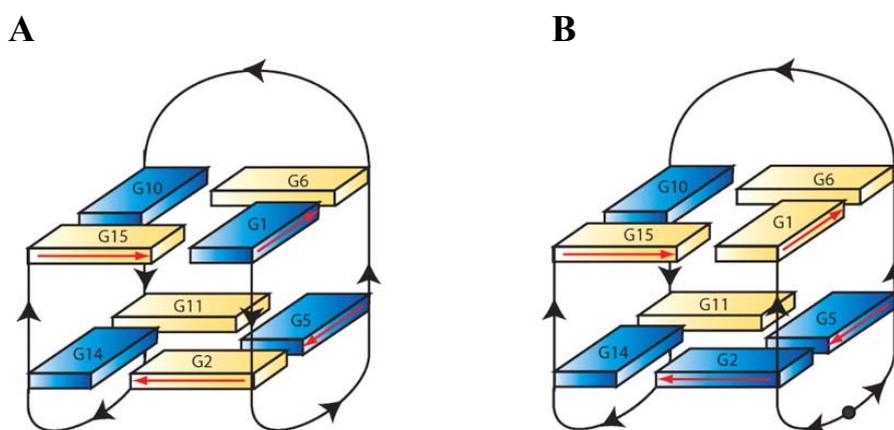
**Fig. 4.1** - *Schematic illustration of the structures adopted by TBA (A) and modified TBA (B). Black arrowheads indicate 5'→3' polarity of the strands. Black circle in mTBA represents the 5'–5' inversion of polarity site. Anti and syn guanines are depicted as yellow and blue solids, respectively. Red arrows indicate the direction of the proton donors and acceptors in Hoogsteen hydrogen bonds.*

4.2 Materials and methods

The oligonucleotides were synthesized as previously reported [106]. The solutions were prepared by dissolving solid lyophilized oligonucleotides in the appropriate buffer and were annealed by heating to 90°C for 5 minutes and slow cooling to room temperature. The buffer used was 10 mM potassium phosphate, 70 mM KCl, 0.1 mM EDTA at pH 7.0. The oligonucleotides concentrations were determined by their adsorption measured at 90°C, in the same buffer, using the same molar extinction coefficient, $\varepsilon(260 \text{ nm}) = 147300 \text{ M}^{-1}\text{cm}^{-1}$. The molar extinction coefficient was calculated by the nearest neighbor model, assuming that the extinction coefficient of the nucleobases involved in the inversion of polarity site, in the modified aptamer, was the same as in the DNA.

Human α-thrombin was purchased from Sigma-Aldrich (product number T6884) and dissolved in the same buffer used for oligonucleotides. Protein solutions were dialyzed against the buffer solution at 4°C by using Spectra Por MW 10.000 membranes, protein concentration was determined spectrophotometrically using a sequence-based extinction coefficient of $66612 \text{ M}^{-1}\text{cm}^{-1}$ at 280 nm.

4.2.1 Differential scanning calorimetry

The total heat required for the unfolding of the TBA and mTBA was measured with a differential scanning calorimeter from Setaram (Micro-DSC III), interfaced with a personal computer. Linear baselines were drawn for each scan, and molar heat capacity values were generated when the molecular weights of the molecules and the molar concentrations used during each scan were introduced. Since $\Delta H_{cal}^{o} = \int \Delta C_p(T) dT$, the area of each peak yields a corresponding transition enthalpy, and the peak maximum yields the transition melting temperature. Given $\Delta S_{cal}^{o} = \int \frac{\Delta C_p(T)}{T} dT$ the calorimetric peak was plotted as $\Delta C_p/T$ versus T to yield a curve, the area of which is $\Delta S°$. The enthalpy and entropy determinations allowed the calculation of the Gibbs free energy via the equation $\Delta G_{298K}^{o} = \Delta H_{cal}^{o} - T\Delta S_{cal}^{o}$. All values reported are an average of at least three repeated measurements.

Two cells, the sample cell containing 800 µL of TBA or mTBA solution and the reference cell filled with the same volume of buffer solution, were heated from 10 to 95 °C at three different heating rates (0.3, 0.5 and 1.0 °C min^{-1}). The molecules concentration range was 8.9 x 10^{-4} - 1.1 x 10^{-3} M. This ramp rate over this temperature range led to no hysteresis and thus the complete reversibility of each scan.

4.2.2 Isothermal titration calorimetry

The binding energetics of both TBA and mTBA to thrombin was obtained with a CSC 4200 Calorimeter from Calorimetry Science Corporation (CSC, Utah) at 25°C. Titration experiments

were done with successive injections of thrombin into the sample cell (1300 µL) containing 1.6 µM TBA or mTBA until saturation was achieved. Titrations were carried out using a 250 µL syringe with stirring at 297 rpm. The solutions of thrombin and DNA samples were prepared in the same buffer. Control experiments were done by injecting thrombin in buffer to obtain the heat effects for dilution of the protein. The calorimetric enthalpy for each injection was calculated after correction for the heat of thrombin dilution. To obtain the thermodynamic properties of interactions a non linear regression analysis was performed using a single set of identical independent binding sites.

4.2.3 Molecular Dynamics simulations

All simulations were performed by means of the GROMACS package [259], employing the all-atom force field parm98 [260]. The initial structures of TBA and mTBA were generated by using the coordinates of the NMR structures (TBA PDB code: 148D, mTBA PDB code: 2IDN), computing the average structure over the 12 and 7 NMR structures, respectively. The mTBA non-standard residue was created using the Xleap module of AMBER 7.0 package [261]. The molecules were neutralized with 14 Na$^+$ ions (placed following electrostatic potential values) and solvated in a cubic box of size 60×60×60 Å^3 containing about 6950 TIP3P water molecules [262]. Initially, water molecules and ions were relaxed by a first steepest descent energy minimization with positional restraints on solute. The LINCS algorithm [263] was used to constrain the bonds and to carry out an initial simulation, 200 ps long, with the positions of the solute atoms restrained by a force constant of 3000 kJ mol^{-1}nm^{-2}, in order to let water molecules diffuse around the complex and to equilibrate. Simulations were carried out with periodic boundary conditions at a constant temperature of 300 K. The Berendsen algorithm was applied for temperature and pressure coupling [264], and the particle mesh Ewald method [265] was used for the calculation of electrostatic contribution to non bonded interactions (grid spacing of 0.12 nm). The trajectory length was 5 ns long for each system.

4.2.4 Docking, complex selection and simulations

To obtain a model of the complex of TBA with two thrombin molecules, rigid-body docking calculations were carried out by using the program ZDOCK 2.3 [122, 266]. The structure of the 1:1 complex between NMR model of TBA and thrombin (hereafter referred to as thrombin A) was kept fixed and, using the TBA molecule as target, a second molecule of thrombin (hereafter referred to as thrombin B) was allowed to rotate and translate around the target. We generated 2000 docked structures, starting from randomly chosen relative orientations. Favorable complexes were selected by a restraint filtering chosen on the basis of the contacts indicated by X-ray data. For TBA the OP atoms of residues G2, G5, T7, T9 have been identified by X-ray to be in contact with the N atoms of side chains of Arg233, Lys236, Lys240 and Arg93 of a second thrombin (the residue numeration is according to X-ray structure [256]). An OP - N (TBA – thrombin B) distance below a cut-off of 6

Å was counted as a restraint satisfaction. These restraints were sufficient for an unambiguous complex selection. Indeed, only one of the 2000 docked complexes was passing the filter and presented the thrombin B in a suitable position for binding. This complex was used as starting structure of MD simulations.

Since in the case of modified aptamer structural data concerning the complex were not available, to obtain a reasonable starting model of the complex of mTBA with two thrombin proteins, the structure of the complex with TBA, obtained by docking, was chosen as template. The backbone of mTBA was superimposed onto that of TBA complexed with thrombin molecules. The resulting mTBA-thrombin complex model was then energy minimized *in vacuo* by 1000 steps of the steepest descent method.

The minimizations and successive MD simulations in explicit water were carried out with the protocol described above. The two complexes, consisting of 9232 proteins atoms and 488 nucleic acid atoms, were solvated in a box of size $85 \times 80 \times 115$ $Å^3$ containing about 22900 TIP3P water molecules. To neutralize the system, 6 water molecules were replaced by Na^+ ions (placed following electrostatic potential values). The trajectory length was 2 ns long for each complex.

4.3 Results

4.3.1 DSC melting study

We have investigated the unfolding of TBA and mTBA using differential scanning microcalorimetry. The DSC melting curves for the two quadruplexes are shown in **Fig. 4.2**, and the corresponding thermodynamic parameters are listed in **Table I**. In the experimental conditions, the transitions of TBA and mTBA are reversible, as demonstrated by the recovery of the original signal by rescanning the same sample. Furthermore, the change of the heating rate from 0.3 to 1.0 °C min^{-1} does not alter the thermodynamic parameters significantly, thereby demonstrating that the studied processes are not kinetically controlled.

The DSC curve of TBA shows a symmetric shape with a maximum centered at 53 °C. The integration of the denaturation peak gives a $\Delta H°(T_m)$ of 96 kJ mol^{-1}. The calorimetric curve of mTBA also shows a symmetric transition and exhibits both higher $T_m = 58$ °C and $\Delta H°(T_m) = 146$ kJ mol^{-1}. The calculated $\Delta G°(298)$ values are 8 kJ mol^{-1} and 17 kJ mol^{-1} for TBA and mTBA, respectively.
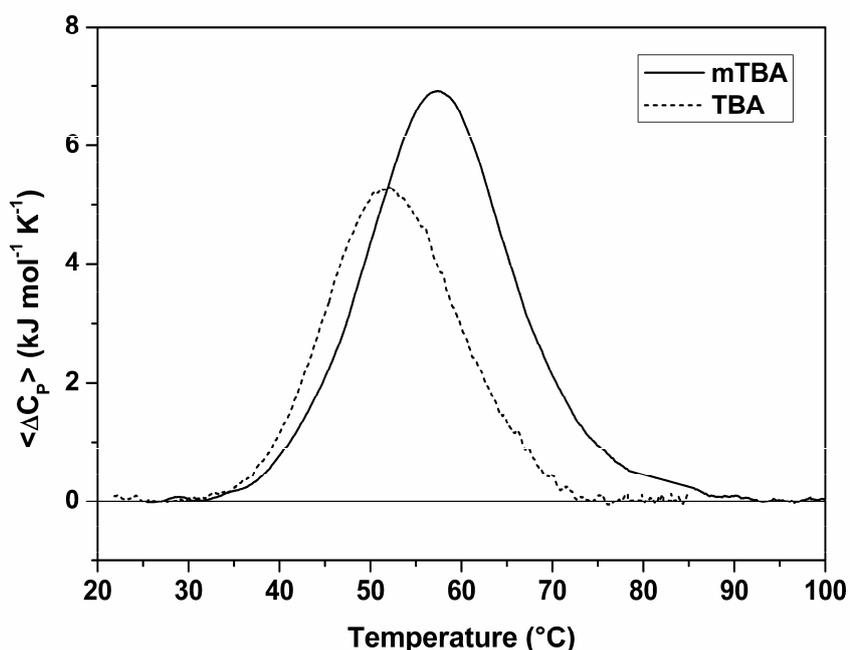
Fig. 4.2 - *DSC profiles for TBA (dashed line) and mTBA (solid line) at 1 °C/min. The experiments were performed in a 10 mM potassium phosphate buffer (pH 7.0) supplemented with 70 mM KCl.*

The thermodynamic parameters were calculated assuming a negligible difference in the heat capacity between the initial and final states, because, for the investigated transitions, the two states have shown comparable heat capacity values. The melting temperatures are in good agreement with

those previously obtained by CD melting profiles (18), whereas the enthalpy values are lower than those derived from the van't Hoff analysis of CD curves, even if the van't Hoff enthalpy calculated for mTBA by DSC curves (163 kJ mol^{-1}) is in good agreement with the experimental value directly obtained by the area under the curves. For the TBA, analogous thermodynamic parameters were also reported by Macaya et al. [249] and recently by Olsen et al. [267] in a calorimetric study of quadruplexes. The latter authors referred the discrepancy between calorimetric and van't Hoff enthalpy values to aggregation phenomena. On the other hand, in order to obtain a calorimetric enthalpy value in agreement with the van't Hoff enthalpy value, few years ago, Smirnov and Shafer [268] analyzed the TBA calorimetric data taking into account a temperature-dependent heat capacities that vary linearly with T.

The whole set of thermodynamic parameters shows that the mTBA is more stable than its unmodified counterpart. The enthalpy value for mTBA results 50 kJ mol^{-1} higher. This result suggests the presence of additional intramolecular interactions. Further, entropy change values suggest that the modified aptamer possesses a more rigid structure respect to TBA.

Table I - Thermodynamic parameters for the denaturation process

	T_m °C	$\Delta H°(T_m)$ kJ mol^{-1}	$\Delta S°(T_m)$ kJ mol^{-1}K^{-1}	$\Delta G°_{298K}$ kJ mol^{-1}
TBA	53.0 ± 0.5	96 ± 5	0.30 ± 0.02	8 ± 1
mTBA	57.9 ± 0.5	146 ± 5	0.43 ± 0.02	17 ± 1

4.3.2 ITC study

To study the thermodynamics of the interaction of TBA and mTBA with thrombin we performed ITC experiments. In the upper panels of **Fig. 4.3** are shown the results of the calorimetric titrations of the aptamers into buffered thrombin solution, conducted at 25 °C. In both cases, an exothermic heat pulse is observed after each injection of protein solution into the aptamer solution. The area of each exothermic peak is integrated and the obtained heat is divided by the moles of thrombin injected. The binding data were corrected for the dilution heats associated with the addition of the thrombin into buffer, separately determined by injection of either thrombin solution into buffer. The resulting values are plotted as a function of the molar ratio, to give the corresponding binding isotherms showed in the lower panels.

The corrected data were fitted using the more simplistic model that assumes a single set of equivalent binding sites to determine the stoichiometry (n), binding constant (K_b), and binding enthalpy ($\Delta_b H°$). The stoichiometry of both complexes clearly indicates that in solution two thrombin molecules bind to one DNA aptamer. Because the two binding modes have not distinct energetic profiles, the application of a two-independent-sites model to fit the titration data provides poorer results. The single set of equivalent sites model is the simplest model to derive K and $\Delta_b H°$. These thermodynamic parameters result averaged on all the microstate values. The binding Gibbs energy change, $\Delta_b G°$, and entropy change, $\Delta_b S°$, are calculated from the equations $\Delta_b G° = -RT \ln K_b$ and $T\Delta_b S° = \Delta_b H° - \Delta_b G°$. The thermodynamic parameters are collected in **Table II**.
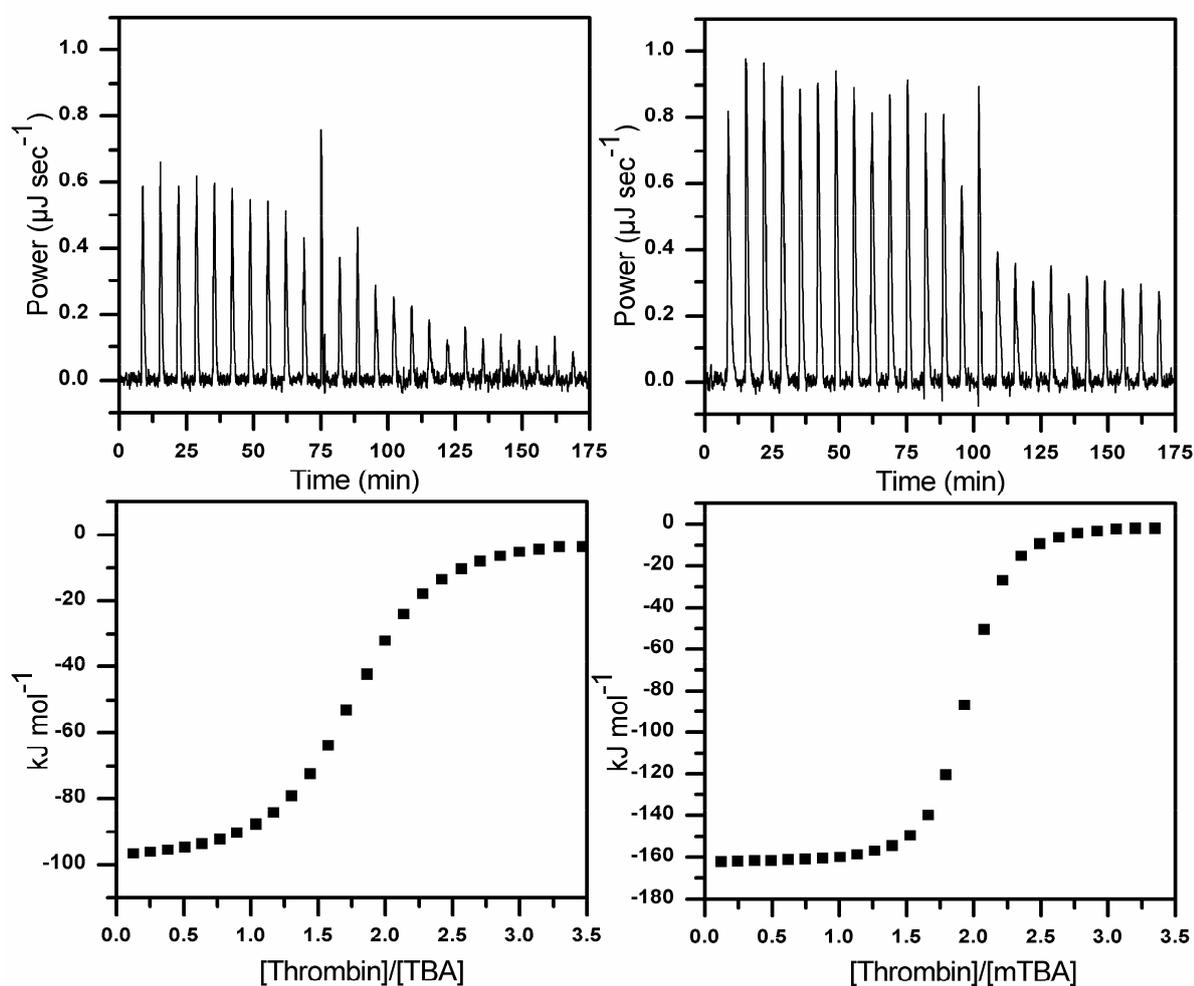

Fig. 4.3 – *Top panels: raw calorimetric data for titration of TBA (on the left) and mTBA (on the right) with serial injections of different volumes of thrombin solution. Bottom panels: binding isotherms resulting from integration of raw calorimetric data after correction for the heat of thrombin dilution.*

The values of the binding constants and the Gibbs energy changes indicate that the associations are strongly favoured, at 25 °C, from thermodynamic point of view. The investigated aptamers bind to thrombin with different affinity. Indeed, the equilibrium constant for the interaction of mTBA with thrombin (K_b = 4 x 10^7 M^{-1}) is about one order of magnitude greater than that for the TBA-thrombin interaction (K_b = 3 x 10^6 M^{-1}). The values of $\Delta_b H°$ and $\Delta_b S°$ show that, in both cases, the binding processes are enthalpically driven; however, the interaction of mTBA with thrombin is associated with a larger favorable enthalpy ($\Delta_b H°$ = -160 kJ mol^{-1}) as compared to TBA($\Delta_b H°$ = -110 kJ mol^{-1}). Our results reinforce X-ray data resolving any doubts on the stoichiometry and indicate that the interaction of the aptamer with two thrombin molecules is not unique to the crystal complex.

Table II - Thermodynamic parameters for the interaction with thrombin

	n	K_b M^{-1}	$\Delta_b H°$ kJ mol^{-1}	$\Delta_b S°$ kJ mol^{-1} K^{-1}	$\Delta_b G°_{298K}$ kJ mol^{-1}
TBA	1.8 ± 0.1	(3 ± 1) x 10^6	-110 ± 9	-0.24 ± 0.03	-37 ± 1
mTBA	1.9 ± 0.1	(4 ± 1) x 10^7	-160 ± 7	-0.39 ± 0.02	-43 ± 1

4.3.3 Molecular Dynamics of TBA and mTBA

MD simulations of the two aptamers produce stable trajectories as shown by macroscopic properties of the systems such as density, potential and total energies (data not shown). The time evolution of the root mean square deviation (RMSD) values, with respect to the initial energy minimized structure, were calculated for all the atoms in the TBA and mTBA quadruplexes, as well as for the guanine base atoms of tetrads (**Fig. 4.4**). It is observed that the all-atoms RMSD gradually increases during the first 1 ns of simulations, and then fluctuate around an average value of 2.1 Å for the TBA and 2.4 Å for the mTBA during the last 4 ns. Interestingly, the RMSD of the guanine bases of tetrads oscillates around an average value of 0.6 Å for the TBA and 0.8 Å for the mTBA, indicating that, in both cases, the G-tetrads are rigid and very stable during the entire simulations.

Stereo diagrams of average structures, over the last 4 ns of MD simulations, are shown in **Fig. 4.5**. The resulting structures are in good agreement with the NMR structures used as starting model. In particular, the unusual characteristics of the mTBA, like the *anti-anti-anti-syn* and *syn–syn-syn-anti* arrangements of the two tetrads, as well as the peculiar groove widths, are maintained in the theoretical structure. Helix parameters of the two average structures were calculated with the

program CURVES [269] (**Tables III** and **IV**). A comparison between the parameters of TBA and mTBA indicates that the base stacking is more efficient in the latter. Indeed, Buckle and Stagger values suggest that the bases within the tetrads of mTBA are more co-planar; and the Rise and Tilt values indicates that the tetrads of mTBA are more parallel to each other in comparison with the tetrads of TBA.
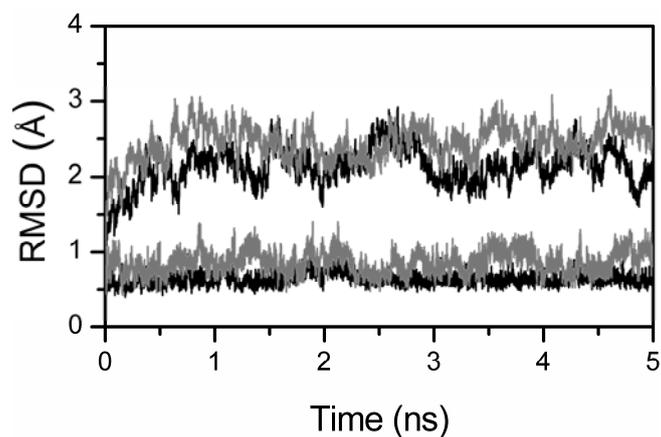

Fig. 4.4 - *Time dependence of RMSD for all the atoms (top plots) and for the guanine base atoms of tetrads (bottom plots) of TBA (black) and mTBA (shaded), during the MD simulations.*
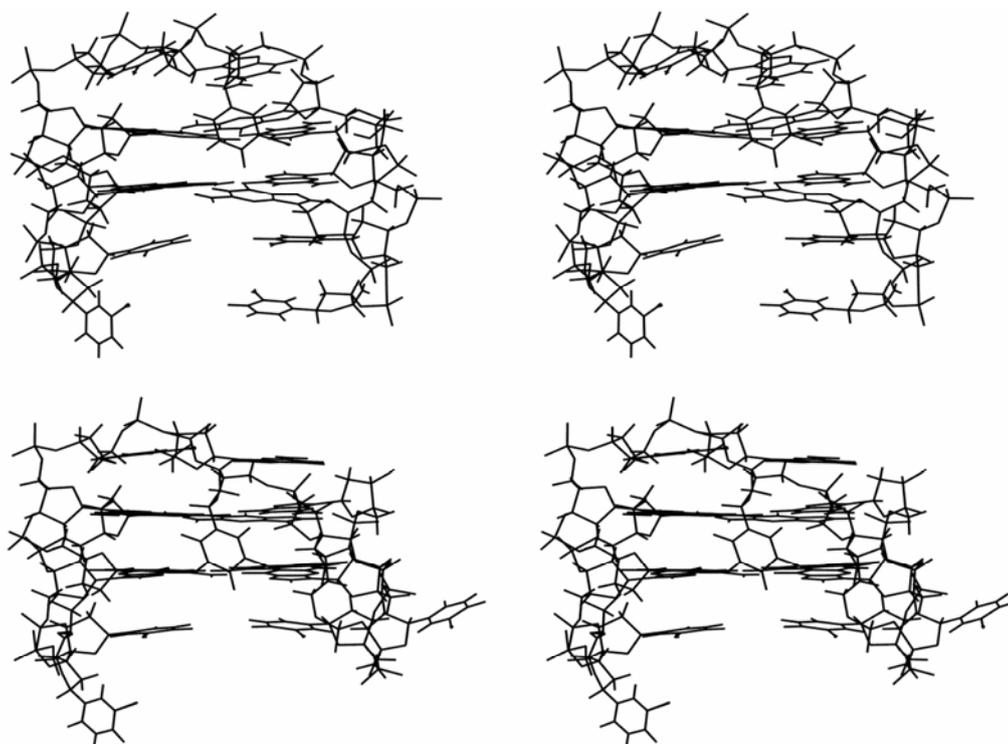
Fig. 4.5 – *Stereoview representation of the average structures, over the last 4 ns of MD simulations, of TBA (top) and mTBA (bottom).*

Table III - Rise, tilt, roll, and twist parameters for the average structures of TBA and mTBA.

TBA	Rise	Tilt	Roll	Twist	mTBA
G1/G2	2.66	-16.96	-177.20	54.22	
	3.37	3.93	175.17	-44.12	**G1/G2**
G5/G6	2.57	14.66	177.85	53.53	
	3.39	-2.09	177.04	-46.06	**G5/G6**
G10/G11	4.01	-5.20	177.24	55.79	
	3.37	-3.10	-178.07	-47.02	**G10/G11**
G14/G15	3.43	2.11	178.40	54.21	
	3.13	-8.38	174.12	-36.78	**G14/G15**

Table IV - Stagger, buckle and propeller parameters for the average structures of TBA and mTBA.

TBA	Stagger	Buckel	Propeller	mTBA
G1/G6	-0.77	15.91	178.84	
	-0.08	0.59	178.09	**G1/G6**
G1/G10	0.26	3.55	-166.37	
	-0.14	10.95	-176.53	**G1/G10**
G1/G15	0.58	16.97	-173.52	
	0.05	4.66	174.25	**G1/G15**
G2/G5	-0.68	-15.74	178.19	
	-0.06	-5.43	-174.12	**G2/G5**
G2/G11	-1.10	15.31	-166.41	
	-0.14	3.92	-173.63	**G2/G11**
G2/G14	-0.19	-2.10	-174.72	
	-0.19	-7.66	-175.04	**G2/G14**

Visual inspection of the trajectories clearly shows that the main differences between the two aptamers concern the behaviour of loops bases. This different behaviour is well emphasized in **Fig. 4.6**, where is reported the difference between TBA and mTBA atomic fluctuations (ΔRMSF). The largest differences in RMSF are associated with the bases T3 (atoms 77-90), T7 (atoms 207-209) and T9 (atoms 272-285) of TBA. The smaller fluctuations of T3 of mTBA in comparison to T3 of TBA could be ascribed to the diverse base orientation, that could confer less mobility. Analysis of the trajectory reveals that the T7 of mTBA, which folds back into a groove, forms long time-residence intramolecular H-bonds with the guanine bases of the tetrads. In particular, three H-bonds were observed: O2(T7)···N2(G6), O4(T7)···N2(G11) and N3(T7)···N3(G11), with H-bond distance < 3.4 Å for more than 80% of the simulated time for the first one, and more than 50% for the last

two. These interactions are not present in the TBA. Visual inspection of the MD simulation also reveals that the T9 of mTBA remains coplanar and parallel to the adjacent G-tetrad, maintaining the stacking interactions during the entire simulation, whereas the T9 of TBA fluctuates towards the solvent.
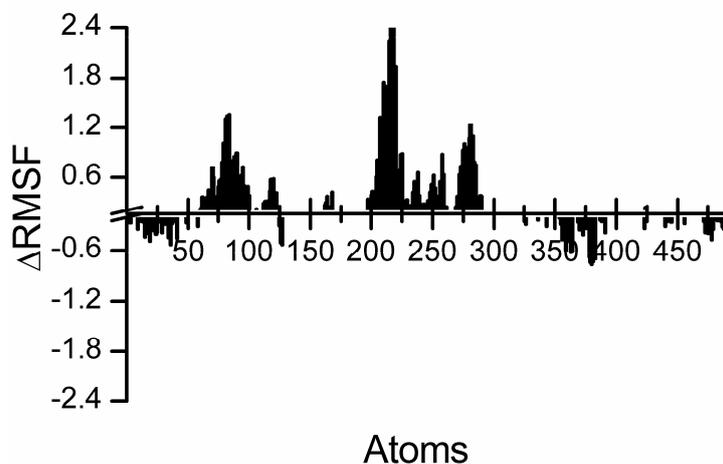
Fig. 4.6 – *Difference between the atomic fluctuations of TBA and mTBA. The largest differences in RMSF (ΔRMSF) are associated with the atoms 77-90 (T3 base), atoms 207-209 (T7 base) and atoms 272-285 (T9 base) of TBA.*

4.3.4 Docking and MD simulations of complexes

We performed ZDOCK calculations with the thrombin molecule and the 1:1 TBA-thrombin complex solved by X-ray. One of the 2000 docked complexes satisfied the filter and presented the thrombin in a suitable binding position (see Materials and methods section). For this complex, 2 ns of MD simulations were performed, and the final structure was used for further analysis. The resulting model of the 2:1 thrombin-TBA complex is reported in **Fig. 4.7A**. The time evolution of the RMSD values with respect to starting structures indicates that the molecular components of the complex have individually reached stable structures during the 2 ns long simulation (**Fig. 4.8A**). In **Table V** the interactions between TBA and (a) the exosite I of thrombin A and (b) exosite II of thrombin B are reported. During the MD simulations a large number of contacts are conserved and additional interactions are formed. In particular, the hydrophobic interactions between T3 and Tyr76 and those between T12 and the residues Ile24, His71, Ile79 and Tyr117 indicated by Padmanabhan and Tulinsky [256], remain unaltered. In contrast, the more weak interaction between T3 and Ile82 is lost. Visual inspection of the complex reveals that additional interactions concerning the exosite I of thrombin A and the nucleotides of the TT loops of TBA are present: (a) the side chain of Arg67 forms an ion pair with O2P atom of T4; (b) the NE and NH2 atoms of Arg75 make

two hydrogen bonds with O2 of T4 and O4 of T13, respectively; (c) the side chain of Glu77 forms a hydrogen bond with nitrogen atom of T12; (d) the side chain of Arg77A forms a hydrogen bond with O2 of T13; (e) the NH_2 group of Asn78 forms a hydrogen bond with phosphate oxygen atom of T13. It is worth noting that His71, Arg75, Tyr76 and Arg77A were reported to be important for inhibition by the thrombin aptamer [251, 253].

Exosite II of thrombin B, as also suggested by Padmanabhan and Tulinsky [256], interacts through numerous ion pairs and hydrogen bonds with TBA residues. In the starting structure, the side chain of Arg233 forms an ion pair with O2P atom of G2. During the simulation this contact is lost and the O1P atom of G2 gains interaction with the side chain of Lys236 which, at the same time, forms another ion pair with O1P atom of G5. The side chains of Lys240 and Arg93 form good ion pairs with O1P atoms of T7 and T9, respectively. These contacts persist for the entire simulation.

MD simulations of the complex of mTBA with two thrombin molecules also produced stable trajectories as shown by RMSD of the components of complex (**Fig. 4.8A**). The structural model of the 2:1 thrombin-mTBA complex is reported in **Fig. 4.7B**. The contacts between mTBA and the exosites of the two proteins are reported in **Table V**. The comparison between the patterns obtained for the complex with TBA and the complex with mTBA indicates that, even though some contacts are conserved, several differences occur. In particular, for the exosite I of thrombin A, residues Ile24, Arg75, Tyr76 and Ile79 make similar interactions with mTBA. On the other hand, the interactions with the side chains of residues His71, Glu77 and Tyr117 are lost in the complex with mTBA. Residues Arg67, Arg77A and Asn78 interact with the modified aptamer, but in a different manner. Indeed, the side chain of Arg67 forms a hydrogen bond with T3 base, the side chain of Arg77A forms a hydrogen bond with O4 of T4, and the NH_2 group of Asn78 forms a hydrogen bond with O2 atom of T12. It is worth noting that Arg75, identified as a key determinant for interaction with TBA [251, 254], makes two identical hydrogen bonds in both models.

Differences occur also in the case of exosite II of thrombin B. Although Arg93 interacts with mTBA in the same manner as with TBA, Lys236 and Lys240 have a different behaviour. Indeed, the side chain of Lys236 makes, simultaneously, two ion pairs with O1P atoms of T4 and G5; and Lys240 forms an ion pair with O1P atom of G6. In addition, the side chain of Arg233 makes two hydrogen bonds with phosphate oxygen atoms of G2. These interactions are not present in the complex with TBA.
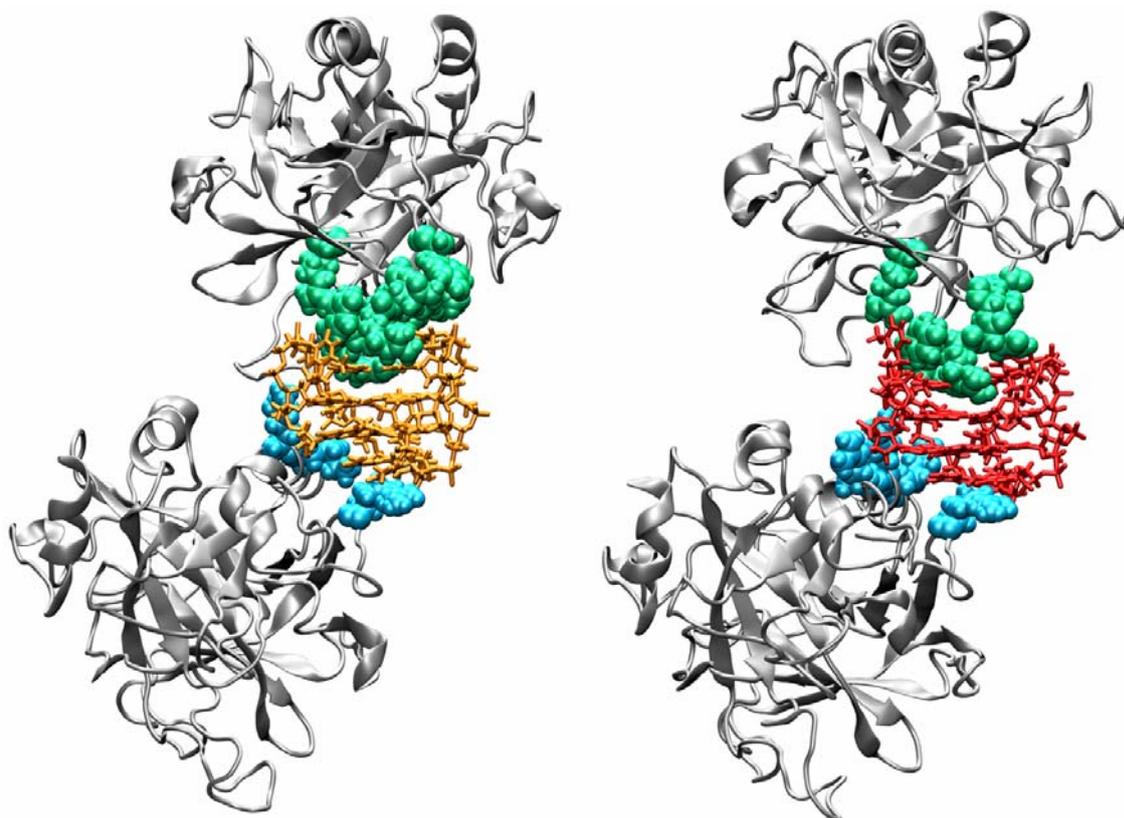
Fig. 4.7 - *Structural models of the complex of (A) TBA or (B) mTBA with two thrombin molecules. The interacting residues of exosite I (green) and exosite II (cyan) are shown.*
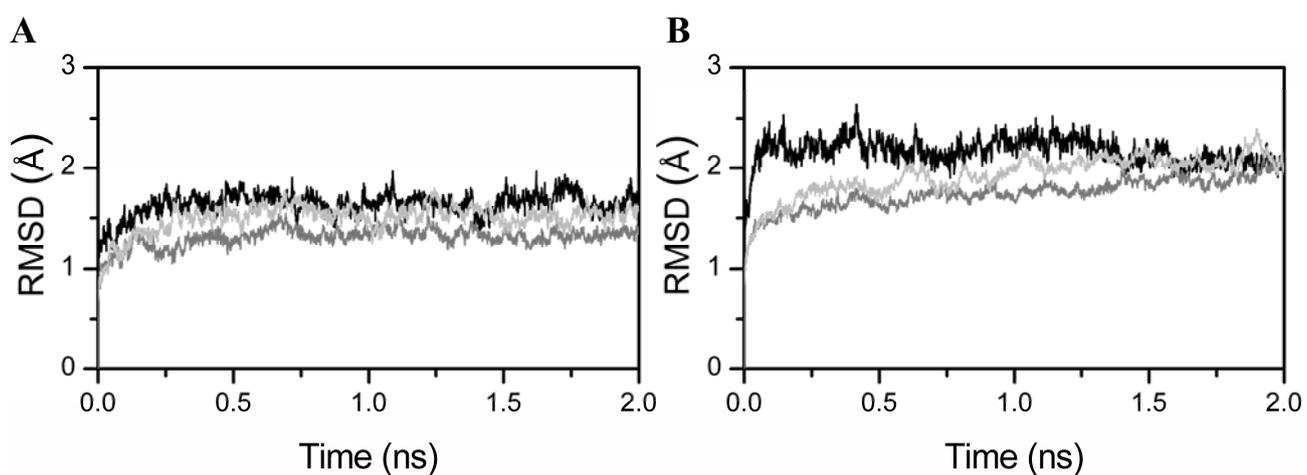
Fig. 4.8 - *Time evolution of RMSD for DNA (black lines), thrombin A (grey lines) and thrombin B (light grey lines) of the complex with TBA (A) and the complex with mTBA (B).*

Table V - Interaction between thrombin and aptamers

(A) TBA

DNA		Thrombin A		Interaction
T3		Tyr76		Hydrophobic
T4	O2P	Arg67	NH1	Ion pair
T4	O2	Arg75	NE	Hydrogen bond
T12		Ile24		Hydrophobic
T12		Ile79		Hydrophobic
T12		His71		Hydrophobic
T12		Tyr117		Hydrophobic
T12	N3	Glu77	OE2	Hydrogen bond
T13	O4	Arg75	NH2	Hydrogen bond
T13	O2	Arg77A	NH2	Hydrogen bond
T13	O2P	Asn78	ND2	Hydrogen bond

DNA		Thrombin B		Interaction
G2	O1P	Lys236	NZ	Ion pair
G5	O1P	Lys236	NZ	Ion pair
T7	O1P	Lys240	NZ	Ion pair
T9	O1P	Arg93	NH1	Ion pair

(B) mTBA

DNA		Thrombin A		Interaction
T3		Tyr76		Hydrophobic
T3	O4	Arg67	NH2	Hydrogen bond
T4	O2	Arg75	NE	Hydrogen bond
T4	O4	Arg77A	NE	Hydrogen bond
T12		Ile24		Hydrophobic
T12		Ile79		Hydrophobic
T12	O2	Asn78	ND2	Hydrogen bond
T13	O4	Arg75	NH2	Hydrogen bond

DNA		Thrombin B		Interaction
G2	O2P	Arg233	NE	Ion pair
G2	O2P	Arg233	NH2	Ion pair
T4	O1P	Lys236	NZ	Ion pair
G5	O1P	Lys236	NZ	Ion pair
G6	O1P	Lys240	NZ	Ion pair
T9	O1P	Arg93	NH1	Ion pair

4.4 Discussion

We have investigated the thermodynamic stability of TBA and mTBA by DSC and the energetics of binding of these molecules to thrombin by ITC. DSC data indicate that mTBA is more stable than TBA. Inspection of **Table I** reveals that the introduction of a 5'-5' inversion of polarity, increases the melting temperature of about 5 °C. The T_m trend reflects the thermodynamic stability as shown by the Gibbs energy values. The enthalpy value for the unmodified aptamer is in agreement with the values previously reported for the same molecule [249, 267]. The $\Delta H°(T_m)$ value for the mTBA results 50 kJ mol^{-1} higher than TBA. These results suggest the presence of additional intramolecular interactions in the modified aptamer. Moreover, the entropy change values suggest that mTBA possesses a more rigid structure respect to TBA. These findings can be rationalized by the results of MD simulations: (a) in the mTBA the T7 forms intramolecular hydrogen bonds with the guanine bases of the tetrads that are entirely lost in TBA; (b) mTBA presents stronger stacking interactions between the G-tetrads as compared to TBA; (c) the T9 base of mTBA, differently from TBA, remains perfectly stacked with the adjacent G-tetrad during the entire simulation. These supplementary interactions of the thymine bases and the strong stacking interactions of the neighboring guanine tetrads may justify the higher enthalpy change observed for the denaturation of mTBA in comparison with TBA. Moreover, the bases T3, T7 and T9 of TBA show larger fluctuations than those of mTBA. The lower mobility of the bases of mTBA loops, in comparison to TBA, suggests that the modified aptamer possesses a more rigid structure with a lower entropy. This could justify the major entropic gain observed for the denaturation process of this molecule.

ITC measurements demonstrate that the binding of TBA and mTBA to thrombin is exothermic in nature and that both aptamers bind with a stoichiometry of 1:2 (aptamer:protein). The latter finding is consistent with the crystallographic studies of TBA complex with thrombin that revealed that TBA was sandwiched between two thrombin molecules, contacting exosite I on one thrombin molecule and exosite II on the other.

ITC data do not permit a distinction between the two sites because the titration data have not distinct energetic profiles, however the thermodynamic data reveals that mTBA has, on average, a higher affinity for thrombin and that its interaction is associated with a larger favorable enthalpy change (**Table II**). The ranking orders of binding constants and enthalpies found in this calorimetric study are not in agreement with the biological activity previously measured *in vitro*, where TBA showed a more marked effect over mTBA [106]. Moreover, Tsiang et al. [253], studying the ability of TBA to inhibit mutant thrombins, have supposed that TBA does not bind to exosite II or that exosite II binding is noninhibitory. Wang et al. [258] additionally showed that the substitution of an

68

abasic residue for either T4 or T13 bases, significantly reduces the inhibition of thrombin. In our structural models, both T4 and T13 bases form hydrogen bonds with residues of exosite I.

The overall picture that emerged from this thermodynamic analysis, taking into account also the findings of other research groups and our MD simulations results, is that the aptamers interact with both exosites I and II, and that the binding to exosite I is the only with inhibitory effect. This could justify not only the binding stoichiometry, but also the disagreement between thermodynamic results and biological activity. In fact, mTBA binds to thrombin with an increased affinity, but this enhancement could be due to a better interaction with the not-inhibiting exosite II.

MD simulations of complexes, which allow an analysis of the binding mode at molecular level, corroborate this hypothesis. Indeed, the molecular models show that the residues of exosite I make more contacts with TBA than mTBA, and, on the contrary, the residues of exosite II make more interactions with the modified aptamer (see **Table V**).

Although mTBA makes less and different interactions with the exosite I of thrombin A, it is interesting to note that the interaction mode of Arg75, identified as fundamental for aptamer binding and inhibition, is preserved also in this case. Other four residues were identified, by mutagenesis study [253], as being important for inhibition by thrombin aptamer: Lys70, His71, Tyr76 and Arg77A. In our models, Tyr76 makes the same electrostatic interaction in both complexes, whereas Arg77A interacts with the modified aptamer in a different manner, and His71 does not make interactions with mTBA. These two differences could explain the decreasing of inhibitory activity. It is also interesting to emphasize that, in both models, the side chain of Lys70 does not interact with the aptamers, but is involved in interactions with other amino acids of exosite I. In particular, Lys70 forms a hydrogen bond with the backbone oxygen atom of Arg75, and two ion pairs with side chains of Glu77 and Glu80 (see **Fig. 4.9**). A mutation of Lys70 breaks these interactions, allowing to the negatively charged side chains of Glu77 and Glu80 to move, preventing the binding of the aptamer.

In conclusion, the stability of TBA and of a modified TBA, containing an inversion of polarity, and the binding energetics of these molecules to thrombin have been investigated experimentally. MD simulations studies of the aptamers have provided information about the different stability behaviour of the molecules and the structural models of complexes with thrombin molecules provided detailed view of the interactions occurring between each aptamer and its target. Biophysical studies here presented show that a fine analysis of the molecular recognition process is fundamental for a rational aptamer design.
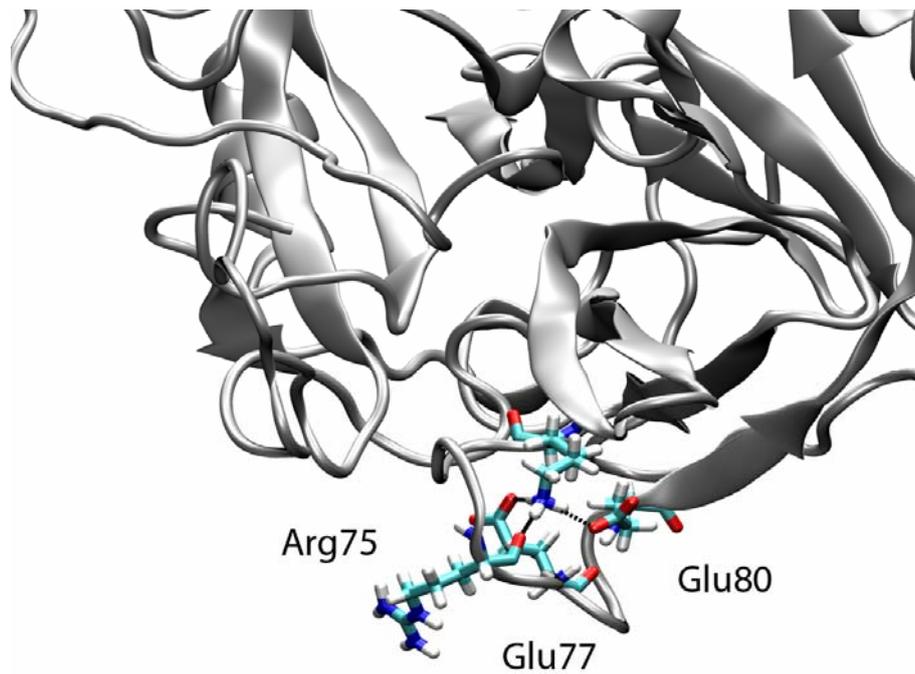
Fig. 4.9 - *Interactions involving the side chain of Lys70 and the residues Arg75, Glu77 and Glu80 of the exosite I*

CHAPTER 5: MOLECULAR DYNAMICS SIMULATIONS OF DNA AND RNA QUADRUPLEX STRUCTURES

5.1 Quadruplex structures of RNA 14-mer r(GGAGGUUUUGGAGG) and DNA 14-mer d(GGAGGTTTTGGAGG)

Recent studies have shown that, similarly to DNA, RNA G-quadruplexes have great biological significance and that there are evidences of RNA quadruplex-forming sequences playing an important role in cellular processes [270]. A crucial example is represented by the fragile X mental retardation protein (FMRP) that binds with high affinity to purine-rich mRNA targets forming quadruplexes, suggesting a key role for these mRNAs in the pathogenesis of the fragile X mental retardation syndrome [271]. Recently, it has been discovered that a short G-rich quadruplex-forming sequence within the 5' untraslated region of the mRNA for the oncogene *NRAS* seems to act as a repressor for translation of mRNA into the protein p21 [272]. Another example of the biological significance of these peculiar RNA structures is modulating gene expression: the gene expression of the Shine-Dalgarno (SD) sequence can be regulated *in vivo*, without altering the GGAGG consensus sequence, by forming RNA quadruplex structures [273].

G-quadruplex structures can fold in different topological arrangements depending on the specific sequence, chain length and presence of monovalent or divalent cations [15]. In some cases, novel non-G-tetrads have been found like the T-tetrad [274], A-tetrad [275], ATAT and GCGC tetrads [276].

Uesugi and coworkers [108] found that the RNA sequence r(GGAGGUUUUGGAGG) folds into an unusual intra-strand parallel quadruplex, containing a canonical G:G:G:G tetrad, a U_4 double-chain reversal loop and a G:G(:A):G:G(:A) hexad. Two r(GGAGGUUUUGGAGG) molecules dimerize in solution by stacking through the hexad-hexad interface (**Fig. 5.1**). The corresponding DNA 14-mer d(GGAGGTTTTGGAGG) is found to form an intermolecular quadruplex with antiparallel strand orientations composed by four canonical G-tetrads and two T_4 diagonal loops (**Fig. 5.1**) [107].

It is rather surprising that a RNA sequence and its corresponding DNA sequence, which differ only for the presence of thymines instead of uracil bases and for deoxyribose sugars instead of ribose, form entirely different quadruplex structures. It is known that RNA sequences prefer to form parallel quadruplexes: the observed difficulty of adopting a *syn* conformation does not promote the formation of quadruplexes containing antiparallel strand orientations. On the other hand. it is also known that DNA sequences can form both parallel and antiparallel quadruplex structures since deoxynucleosides can adopt easily both *anti* and *syn* glycosidic bond conformation.
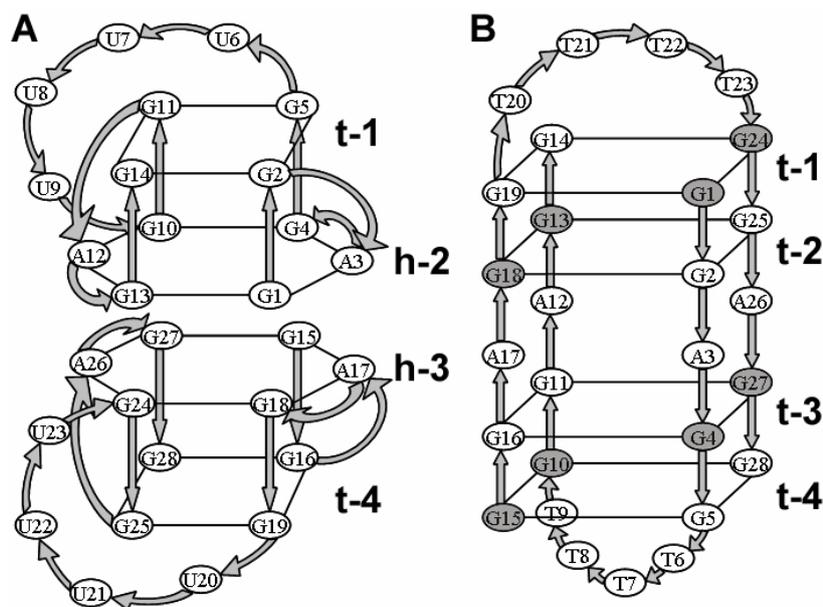
Fig. 5.1 – *Schematic representation of the quadruplex structures formed by (A) r(GGAGGUUUUGGAGG) and (B) d(GGAGGTTTTGGAGG). Rectangles indicate the G-tetrads (planes t-n) and hexagons the hexads (planes h-n).*

Interestingly, the adenine residues of the DNA quadruplex (D14) are located at the center of the core helix, between two G-tetrads, whereas in the RNA quadruplex (R14) two adenine residues associate with a G-tetrad forming a hexad, creating in such way a larger surface that acts as dimerization surface (**Fig. 5.2**).

Considering that GGA triplet repeats have been found in biologically significant parts of DNA [277] and RNA [278] and that the GGAGG sequence has been shown to block the HIV-1 RNA dimerization *in vitro* [279], it is of particular interest to study in details the folding topology of these sequences. So far, there are relatively little experimental data on RNA quadruplexes and, in particular, the folding of RNA into quadruplex structures is largely unexplored.

In order to improve the understanding of the main factors contributing to the stability of this particular fold of RNA, and to investigate the role of adenines in the different topologies of the D14 and R14 quadruplexes, we have performed extensive Molecular Dynamics (MD) simulations of these molecules in explicit solvent in the presence of Na$^+$ ions. Moreover, to investigate the ability of the d(GGAGGTTTTGGAGG) sequence to form, as in the case of RNA, a stable dimeric quadruplex architecture comprised of two hexads and two tetrads, we constructed in silico a chimeric molecule (hereafter referred to as R-D14) with the DNA sequence of D14 and the structure of R14, that was also subjected to MD simulation.
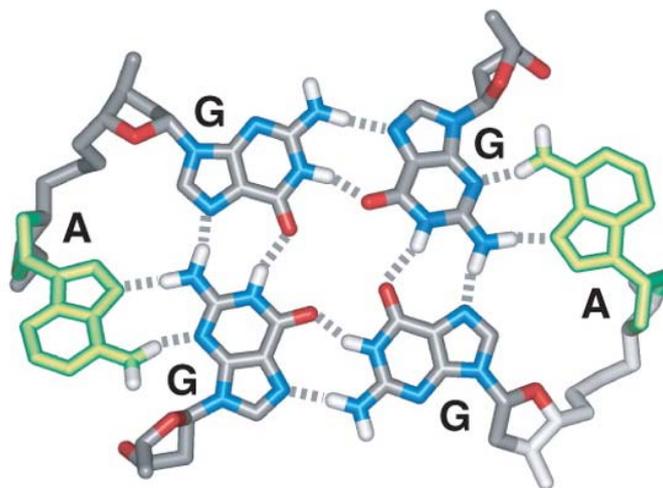
Fig. 5.2 - *Top view of a hexad plane*

In recent years, computational studies have proved particularly useful in characterizing a wide range of structures of nucleic acids and their binding modes to cations [280-282], but so far, no extended MD simulations studies were performed on RNA hexads. Moreover, no evidence at atomic detail is available about the coordination of these structures to cations, and, in particular, about the coordinating role of the hexads. We elucidate here a possible mechanism of coordination to Na^+ ions.

5.2 Methods

All simulations were performed by using the GROMACS package [259], employing the all-atom force field parm98 [260]. The initial structures of the R14 and D14 were generated by using the coordinates of the previously reported structures [107, 108]. The chimeric R-D14 molecule was generated in silico from the R14 structure by replacing the uracil bases with thymine bases and the ribose sugars with the deoxyribose sugars. The initial structure of R-D14 was energy minimized *in vacuo* by 1000 steps of the steepest descent method. The molecules were then neutralized with 26 Na^+ ions (placed following electrostatic potential values) and solvated in boxes with more than 5000 TIP3P water molecules [262]. Initially, water molecules and ions were relaxed by energy minimization and allowed to equilibrate for 200 ps of MD at 300 K with the solute molecules restrained at their initial geometry. The bonds were constrained by the LINCS [263] algorithm by means of a force constant of 3000 kJ mol^{-1} nm^{-2}. Finally, the equilibrated systems were subjected to unrestrained MD simulations for 40 ns. Simulations were carried out with periodic boundary conditions at a constant temperature of 300 K. The Berendsen algorithm was applied for temperature and pressure coupling [264], and the particle mesh Ewald method [283] was used for the calculation of electrostatic contribution to non bonded interactions. MD trajectories were analyzed using GROMACS analysis tools.

5.3 Results

Three independent MD simulations have been performed in the course of this study. The first two simulations were carried out on the quadruplex structures formed by two molecules of RNA 14-mer r(GGAGGUUUUGGAGG) (R14) and by two molecules of DNA 14-mer d(GGAGGTTTTGGAGG) (D14), whereas a third simulation was performed for a chimeric molecule with the DNA sequence of D14 and the structure of R14 (R-D14). The length of each simulation was 40 ns for a total of 120 ns.

D14 and R14 simulations were started from the high-resolution NMR structures previously published [107, 108], while the starting structure of R-D14 was generated from the R14 structure. Since no evidence at atomic detail is available about the cations coordination of these structures, the molecules were immersed in water solution with the Na^+ ions not directly coordinated to the molecules. The binding mode of the cations to the quadruplex molecules was then revealed by monitoring the spontaneous insertion of the ions in the central channel of the structures under study. To observe whether extensive MD simulations lead to these structures being disrupted, ions being expelled, or more ions replacing the water molecules in the central channel of the quadruplex, we run the simulations monitoring them until 40 ns for each molecule.

5.3.1 MD simulation of R14

MD simulation of R14 produces a stable trajectory, the root mean square deviation (RMSD) value of the theoretical structure as compared to the starting NMR structure is 2.4 ± 0.1 Å. The final structure of R14 along with the coordinated cations after 40 ns of MD simulation is reported in **Fig. 5.3**. The final coordination shows two Na^+ ions coordinated at the carbonyl oxygens of the guanine bases, positioned between the first tetrad (t-1) and hexad (h-2) planes, and the second hexad (h-3) and tetrad (t-4) planes, respectively.

Analyzing the simulation we observe that the central channel of R14, initially without any coordinated ions, was hydrated within the first few picoseconds by the entry of water molecules. After less than 100 ps of simulation one Na^+ ion enters the channel from the bottom and situates itself between planes h-3 and t-4. After 1 ns another cation moves to the top of the channel, coordinating the 4 carbonyl oxygens of the G-tetrad forming t-1, positioned almost coplanarly to the tetrad. This situation remains unperturbed for about 500 ps of simulation (from 1 to 1.5 ns absolute simulated time), then the cation enters the channel displacing one water molecule and positioning between planes t-1 and h-2. No changes in the coordination of cations were observed until end of trajectory, but only small movements of the cations inside the coordination site. Indeed, during the remaining simulation the two Na^+ ions are temporarily positioned exactly at the midpoints between

hexad and tetrad planes, while sometimes are more close to the hexad plane, never leaving the coordination site.

Interestingly, no ions occupy the quadruplex channel between the planes of the hexads, but a water molecule occupies this position after few ps of simulation and remains trapped at this site for the entire simulation.

It is worth noting that the movement of ions from solvent into the central channel has a marked effect on the quadruplex structure and stability. In fact, in the absence of any coordinated cation, the guanine bases initially tend to rotate, slightly deforming the canonical G-tetrad scheme in which the four guanines are associated through a cyclic array of eight hydrogen bonds. Additionally, without any coordinated cation, the two outer G-tetrads at top and bottom of the molecule give rise to a bifurcated hydrogen bond geometry involving N1, N7, and O6 atoms of guanines, as already seen in other MD simulations of quadruplex molecules, but for inner G-tetrads [284]. On the contrary, bifurcated geometry was never observed in the simulation of R14 for the inner guanines forming hexads. Probably this is due to the expansion of the G-tetrads into the hydrogen-bonded hexads, conferring more rigidity to the guanine bases and not allowing them to shift.

All glycosidic bonds of guanine and uracil residues in the starting structure of R14 are in *anti* conformation, while the adenine residues are in a high *anti* conformation [108]. Throughout the trajectory, all glycosidic bond conformations of guanine and adenine residues are conserved, conversely, two uracil residues of the loops (U9 and U21) adopt a *syn* glycosidic bond conformation. This happens because the two double-chain reversal loops of R14 are very flexible regions. This behaviour is described in **Fig. 5.4**, where the atomic fluctuations (RMSF) of selected R14 residues clearly show the higher fluctuation of uracil residues compared to the rest of the molecule. Visual inspection of the trajectory reveals that the uracil basis are not coplanar and hydrogen bonds are not formed between them. In addition, no uracil bases are stacked on the closest G-tetrads, except for the first residues of each loop (U6 and U20) that are sporadically stacked on the adjacent guanine residues.

During the simulation, the two adenine bases of each hexad are hydrogen-bonded to guanines and stacked to the other two adenines at the dimerization interface. In addition, all the four adenine bases are slightly rotated out of plane of the guanines to form a hydrogen bond between the NH_2 group of adenines and the O4' atom of the adjacent guanosines.

As previously mentioned, binding to Na^+ leads to a stiffening of the molecule, reorientation of the bases and consequent stabilization of the molecule. Indeed, the presence of cations in the quadruplex channel minimize the electrostatic repulsion between the O6 atoms of the tetrad and hexad planes that, consequently, move closer to each other, as indicated by the distances of the two

planes' centers of mass reported in **Fig. 5.5**, optimizing the stacking interactions. This stabilization is also reflected in the average number of hydrogen bonds between the bases of each plane of R14 that, after ions coordination, stabilizes around an average value of 8 for the tetrads and 12 for the hexads.

The previous observations are summarized by the energy contributions as calculated over the first and last 200 ps of the trajectory. Especially interesting is the comparison between the van der Waals energy (that comprise the stacking contributions) of the molecule that from a starting value of -2150 (± 45) kJ/mol decrease to a more favorable -2370 (± 40) kJ/mol. Likewise, the RNA-Na$^+$ electrostatic energy contribution goes from -760 (± 80) kJ/mol to a value of -1100 (± 100) kJ/mol, contributing to the stabilization of the system.

A **B** **C**
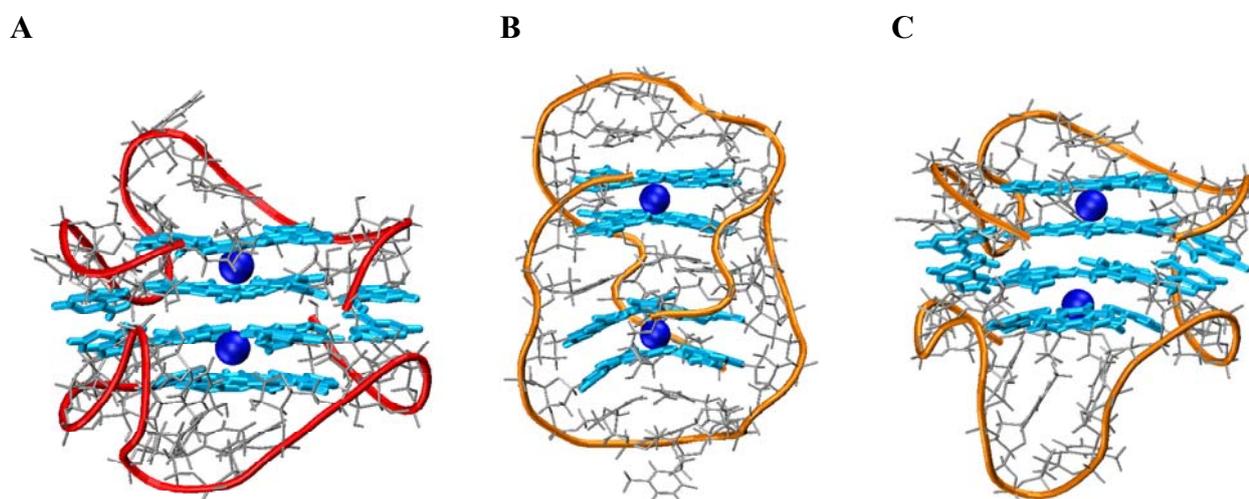
Fig. 5.3 - *Snapshots of the final structures along with the coordinated cations after 40 ns of simulation for R14 (A), D14 (B) and R-D14 (C).*
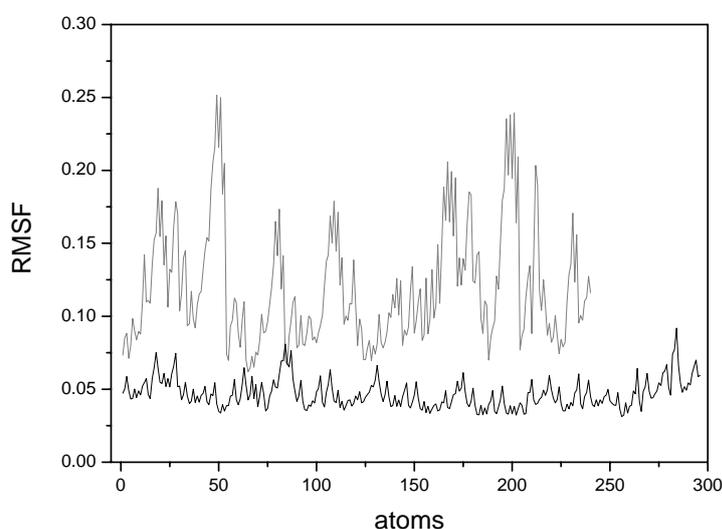
Fig. 5.4 - *Atomic fluctuations of uracil residues (shaded line) and of guanine and adenine residues (black line) of R14.*
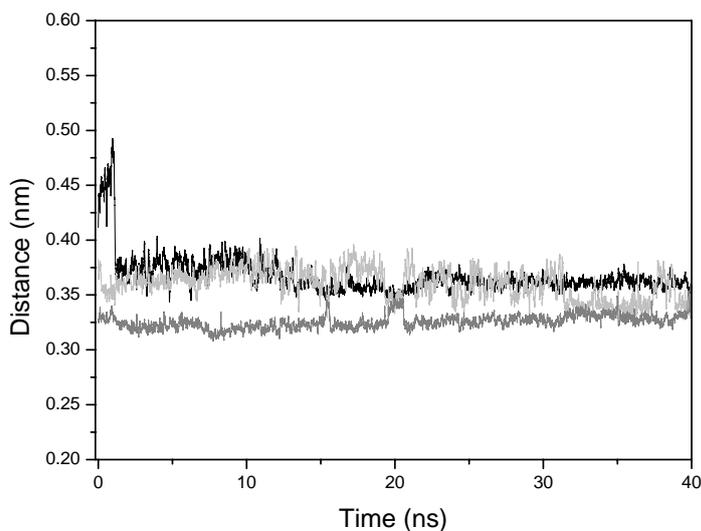
Fig. 5.5 - *Distances between the centers of mass of planes t-1 and h-2 (black), h-2 and h-3 (grey), h-3 and t-4 (light grey).*

5.3.2 MD simulation of D14

MD simulation of D14 also produces a stable trajectory, resulted in a RMSD value along the trajectory between starting NMR and theoretical structures of 2.7 ± 0.1 Å. The final structure of D14 along with the coordinated Na^+ ions after 40 ns of MD simulation is shown in **Fig. 5.3**. D14 coordinates a total of two Na^+ ions placed between the first (t-1) and second (t-2) tetrad planes, and between the third (t-3) and fourth (t-4) tetrad planes, respectively.

As for the case of R14, the central channel formed by the tetrads of D14, initially without any coordinated ions, was fully hydrated within the first few picoseconds by the entry of water molecules. Especially hydrated was the cavity formed between planes t-2 and t-3 where the adenines leave space to the water molecules. In fact, the adenine bases of D14 clearly do not form an A-tetrad during the MD simulation and are relatively mobile. The time evolution of the number of hydrogen bonds formed between adenines fluctuates around an average value of 1, indicating the formation of a single interaction that, in particular, is the hydrogen bond formed between the N3 atom of A12 and the NH_2 group of A17.

The binding process of the Na^+ ions to D14 during the simulation occurs on slower time scales than for R14, in fact the first cation takes position at midpoint between planes t-3 and t-4 after 7.5 ns of simulation, while the second cation situates itself between planes t-1 and t-2 after 32 ns. This is probably due to the presence of the diagonal loops that obstruct the channel, not allowing ions to enter from the top and the bottom of the molecule. Indeed, both the first and second ions coordinating the quadruplex pass through the central cavity between planes t-2 and t-3 that, thanks

78

to the flexibility exhibited from the adenines, results quite accessible to the ions. This cavity is not a stable binding site for the cations, which are only temporarily coordinated during the simulation.

After about 2.5 ns of simulated time one Na$^+$ ion enters the central cavity and situates itself between tetrads t-3 and t-4, but closer to t-3 and coordinating its four carbonyl oxygens. This situation remains unchanged until 7.5 ns, then the cation moves downwards and binds exactly at the midpoint between planes. After 22.5 ns of simulation another cation situates itself in the central cavity and coordinates from bottom the carbonyl oxygens of the tetrad forming t-2, in position almost coplanar with the tetrad. Then, at about 28 ns of simulation the cation moves between tetrads t-1 and t-2, but closer to t-2, and finally at 32 ns situates itself at the midpoint between tetrads. This situation remains unchanged for the rest of the simulation.

Visual inspection of the trajectory before the binding of cations reveals that the guanine bases, as in R14, are subject to an in-plane rotation that slightly distorts the G-tetrads. Although complete ions coordination was fully achieved only after 32 ns, the molecule remains rather stable prior to this. Nevertheless, the coordination of cations has a marked stabilizing effect on the quadruplex, indeed, after Na$^+$ coordination occurs, the guanines do not rotate anymore. Additionally, the average number of hydrogen bonds between the guanine bases forming the tetrads remains fixed around an average value of 8, while before coordination this value was fluctuating around an average of 5-6 per tetrad.

Throughout the simulation, the *syn-syn-anti-anti* glycosidic bond conformation of the guanine bases in the tetrads is well conserved, as the *anti* glycosidic bond conformation of thymine residues in the two diagonal loops.

Both loops show a significant low flexibility during the simulations. The first and third thymine residues of each loop are positioned on top of the outer G-tetrads (t-1 and t-4) of the quadruplex and parallel to them, forming stacking interactions with the aromatic rings of the guanine bases. During the simulation, we observed these thymines forming a hydrogen bond between the O2 oxygen and the N3 hydrogen atoms of the first and third thymine residues, with an average donor-acceptor distance of 2.8 Å. The hydrogen bond between the first and third thymine residues of the bottom loop (T6 and T8) remains stable during the entire simulation. Conversely, in the top loop, the first residue of the loop (T20) loses the interaction with the third residue (T22) after 15 ns of simulation and, simultaneously, forms a hydrogen bond with the fourth thymine (T23). This hydrogen bond, formed between the N3 hydrogen atom of T20 and the O4 oxygen of T23, remains stable until end of simulation, with a N3-O4 distance of about 2.8 Å. Moreover, the T22 residue continues to be close to the other two thymines and maintains its coplanar arrangement, stacked on top of the adjacent guanine residues, although it is no longer hydrogen bonded to T20 (**Fig. 5.6**).
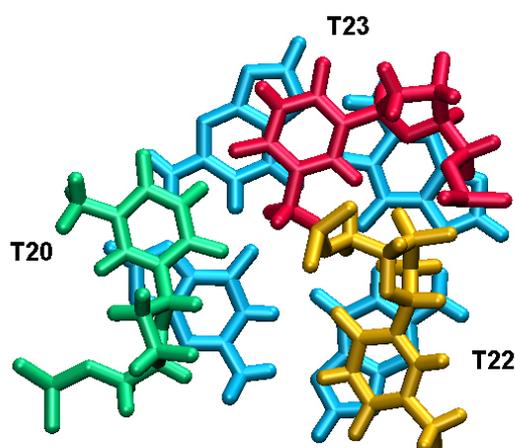
Fig. 5.6 – *Thymines arrangement in the top loop of D14.*

As mentioned above, the adenines are quite flexible and two out of four residues (A3 and A26) change the glycosidic bond conformation, while the other two residues (A12 and A17), that are more rigid because involved in a hydrogen bond, conserve the glycosidic bond conformation throughout the trajectory. In particular, the starting *anti* glycosidic bond conformations of A12 and A17 are conserved, conversely, the A3 converts from *anti* to *syn* conformation, while the A26 converts from *syn* to *anti* conformation.

The energy contributions, as calculated over the first and last 200 ps of the D14 trajectory, reveal that of particular relevance is, in this case, the DNA-Na$^+$ electrostatic energy contribution that from an initial -340 (± 60) kJ/mol value becomes -1200 (± 80) kJ/mol in the final structure.

5.3.3 MD simulation of R-D14

The R-D14 simulation stabilised to a RMSD value of 2.7 ± 0.1 Å with respect to the starting structure. At the end of MD simulation, the R-D14 molecule shows the same Na$^+$ coordination pattern as the R14 quadruplex (**Fig. 5.3**). Analyzing the trajectory, we observed that after about 200 ps of simulation the first Na$^+$ ion moves from the solvent to the channel, positioning itself between h-3 and t-4. Next, no other ions coordinate the molecule until 24.5 ns, when a second cation enters the channel placing itself between t-1 and h-2. The absence of any cation coordinated to the bases forming t-1 and h-2 for 24.5 ns causes a large motion of these residues. In particular, during the simulation the guanine bases of t-1 rotate out of plane, partially disrupting the cyclic array of hydrogen bonds of the tetrad. In addition, the movement of the adenine bases of h-2, involving also the adenines of the stacked h-3, causes the breaking of a number of hydrogen bonds between adenine and guanine bases of both hexads. On the other hand, the guanine bases of the core of h-2 remain relatively stable, probably thanks to the stacking interactions with the closer guanines of h-3.

Despite these marked fluctuation, the R-D14 molecule does not dissociate and the stacking between guanines as well as between adenines at the dimer interface of the two hexads is maintained for all the simulated time. The coordination of the second cation leads the guanine bases of t-1 to return coplanar each other and parallel to the h-2 plane, restoring the hydrogen bonds of the tetrad. In contrast, since the adenines do not return relatively coplanar to the plane of the guanines of the hexad, one of the two hydrogen bonds between adenine and guanine bases of each hexad and the hydrogen bond between the NH_2 group of adenines and the O4' atom of the close guanosines are not restored.

The *anti* glycosidic bond conformations of all guanine and adenine residues are conserved throughout the trajectory. On the contrary, two thymine residues of the loops (T7 and T22) convert the glycosidic bond conformation from *anti* to *syn*.

Concerning the loops behavior, visual inspection of the simulation reveals that the thymine bases do not form hydrogen bonds between them and that no thymines are stacked on the G-tetrads. At the same time the last two thymines of each loop, although turned toward the solvent, form stacking interactions.

5.4 Discussion

We have carried out extensive MD simulations of quadruplex structures formed by two oligomers of RNA 14-mer or DNA 14-mer in explicit water and in the presence of Na^+ ions in solution.

The focus of our study was the dynamic behaviour of these quadruplex molecules, the binding mode of the ions, and their influence in stabilizing these quadruplex structures, since no evidence is available until now about the coordination of ions to these structures. We also aim to improve the understanding of the main factors contributing to the stability of the unusual folding of R14 and, in particular, about the role of adenine bases.

Both R14 and D14 quadruplexes show a final coordination of two Na^+ ions coordinated in the central channel at the carbonyl oxygens of the guanine bases. Starting without any Na^+ directly coordinated to the quadruplexes, the simulations reveal that in both cases the cations spontaneously move into the central channel of the structures under study, without causing dissociation of the quadruplexes and with a marked stabilizing effect. Interestingly, the coordination process to the two quadruplexes is rather different, indeed, for the R14 molecule, the Na^+ ions enter the channel from the top and the bottom of the quadruplex, leading to a faster process (it occur within the first 2 ns), whereas, to coordinate the D14, the cations go through a cavity sited between the inner tetrads, leading to a slower coordination process (coordination was achieved at 32 ns).

Brownian dynamics simulation for the sequence d(GGAGGAG), which folding is comparable to R14 (but with a hexad motif composed of two different strands), have predicted five potential ion-binding sites, three of them in the channel between the four planes, and other two in the grooves between hexads [285]. Accordingly to these predictions, we found two Na^+ ions sandwiched between the planes of R14, conversely, we do not found Na^+ ions between hexads and do not observe in the grooves between hexads of R14 quadruplex any electronegative pocket necessary for the binding of cations.

The simulation of R14 shows low fluctuations for the residues forming tetrads as well for those forming hexads, especially after cations coordination. In particular, the position of the adenine bases in the hexads and their hydrogen bonding pattern are well conserved. On the other hand, high fluctuations were observed for the residues of the two double-chain reversal loops.

D14 simulation also shows low fluctuations for the guanine residues of the tetrads, whereas high fluctuations were observed for the adenine bases, which are relatively free to move. Conversely, the positions of the thymines in the two diagonal loops are well conserved, revealing a stable nature of these loops.

Concerning the R-D14 simulation, some considerations need to be done. We constructed the R-D14 molecule in order to investigate the ability of the DNA sequence d(GGAGGTTTTGGAGG) to

form, as in the case of RNA, a stable dimeric quadruplex architecture comprised of two hexads and two tetrads. Patel and coworkers found that the DNA sequence d(GGAGGAG) in Na$^+$ solution is able to form a G:G(:A):G:G(:A) hexad motif composed of two different strands, also in that case stabilized by the stacking of the two hexads [285]. Moreover, circular dichroism (CD) experiments revealed that at low concentration of cations in solution, the sequence d(GGAGGTTTTGGAGG) shows a similar spectral pattern as r(GGAGGUUUUGGAGG) with a positive band at 262 nm [286], characteristic of a parallel quadruplex structure, suggesting or at least not excluding the possibility that, under particular conditions, this DNA sequence could folds as the R14 molecule. In silico experiments like this one can help understanding the main factors contributing to the stability of different structural elements in determining a particular fold.

The MD simulation of R-D14 shows that d(GGAGGTTTTGGAGG) DNA is capable of forming cation-stabilized parallel quadruplexes with the same folding of R14. However, the simulation of this hypothetical structure revealed significant differences between the DNA and RNA quadruplexes. Throughout the simulation, the R-D14 molecule appears to be more unstable than the other two quadruplexes when the central channel is empty, although the quadruplex structure is maintained. Moreover, the final cation-stabilized structure shows a partial rupture of the hydrogen bonding pattern of the hexads.

The main difference between the RNA and DNA molecules studied here is to be found in the presence of the hydroxyl groups in the sugar moieties of RNA. In fact, the 2'-OH groups in the RNA promote the formation of a large number of intramolecular hydrogen bonds that contribute greatly to the stability of the R14 molecule. The total number of intramolecular hydrogen bonds formed by R14 is on average 53, while this number is 38 and 37 for D14 and R-D14, respectively. As expected, we observed long time-residence intramolecular hydrogen bonds between 2'-OH and any of the hydrogen-bond acceptor atoms of the next neighbouring nucleotide. From our analysis, the hydrogen bonds with the longest time-residence are the ones formed with phosphate oxygen atoms (**Fig. 5.7**), and in particular, the hydrogen bonds 2'-OH(G1)···OP(G2), 2'-OH(G10)···OP(G11), 2'-OH(G15)···OP(G16), and 2'-OH(G24)···OP(G25).

These results suggest that the nucleotides G1-G2-A3 and G10-G11-A12 (and their symmetrical counterparts for the second monomer) forming the 'sharp V-shape turn' that allows for the hexad formation [108], are strongly stabilized by the particular network of hydrogen bonds formed around the adenines, as shown in **Fig. 5.7**. This network of hydrogen bonds create a robust scaffold which allows for more stability of the molecular architecture and that infers more rigidity to the stacked bases. In particular, the hydrogen bond formed by the 2'-OH with the phosphate oxygen acts as a 'clip', supporting the geometry of the V-turn.

The loss of hydroxyl groups in the sugar moieties of DNA implies a loss of several hydrogen bonds, however, the results show that d(GGAGGTTTTGGAGG) can folds as the r(GGAGGUUUUGGAGG) even though the structure is clearly less stable.

It is reasonable to suppose that, during the quadruplex formation, some of these d(GGAGGTTTTGGAGG) fold as the R14 and then convert their structure to a more stable antiparallel structure. In this view, the stable nature of the loops in the antiparallel structure could have a key role in determining the folding topology.
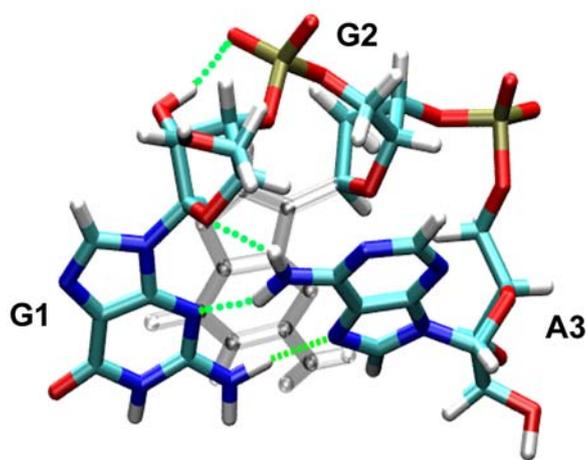

Fig. 5.7 - *Network of hydrogen bonds formed around the adenine residues of R14 quadruplex.*

CONCLUSIONS AND PERSPECTIVES

In recent years telomeres and telomerase have attracted the interest of investigators, due to their involvement with aging and cancer. The telomere, as the guardian of the chromosome, performs essential functions that are regulated as part of the cell cycle, although many details of the events and players in telomere maintenance are as yet unclear. The only certainty is that telomerase, the key enzyme for telomere lengthening, is overexpressed in about the 85% of cancer cells, thus favouring their immortalization. Hence, telomeres and telomerase represent promising targets for anti-cancer therapy. Among the strategies to target telomeres and telomerase, G-quadruplex-interactive compounds appear to provide promising approaches. Several classes of molecules have been found to stabilize the folding of G-rich telomere strands into G-quadruplex structures, thereby inhibiting telomerase access. The ultimate aim of much of this work is to develop compounds that have the ability to target quadruplex DNAs at specified base sequences or structural motifs. The obvious benefit of this research is that we would create novel therapeutic agents with greater efficacy and less toxicity than existing compounds used in the clinic, particularly anticancer, antibiotic and antiviral drugs. It is important that investigators aim at a deeper understanding of the structural aspects of human telomere folding and the energetic aspects of quadruplex-drug molecular recognition. Many classic or newly adopted physico-chemical techniques can be very useful towards this direction, even if ITC is clearly the method of choice to measure the free energy, enthalpy, and stoichiometry of a quadruplex-drug binding reaction.

The data reported in this thesis demonstrate that the interaction between drugs and DNA quadruplex structures is regulated by many factors. In particular, cation-induced rearrangements and any structural factor involving accessibility of drugs to the quadruplexes are important. Since grooves size and their accessibility vary in relation to the quadruplex topology [13], the study of groove binders could offer the opportunity to increase selectivity for a specific quadruplex structure. This encourages the design and the investigation of new quadruplex ligands that could be capable to discriminate among several quadruplex scaffolds.

Further, the therapeutic potential of quadruplex oligonucleotides as aptamers has been shown in the last years. Aptamers based on the quadruplex motif may prove useful for binding and inhibiting specific proteins or even as pharmaceutically active agents. The contributions of biophysical studies to our understanding of nucleic acid folding principles will contribute to the rational design of nucleic acid aptamers for a variety biomedical and other applications. Moreover, as shown in this thesis, the incorporation of modified residues into oligonucleotides could affect their folding and produce useful changes in biophysical properties of the resulting quadruplexes.

Taken altogether, the available data point to important roles for DNA quadruplexes in biological systems and argue strongly for further studies of their sequence-dependent structure and stability. So far, a question that remains to be fully answered concerns the folding of RNA into quadruplex structures. This field is largely unexplored, and it would be unsurprising if these RNA quadruplexes had greater complexity, given, for example, the propensity of RNA molecules to use the 2'-hydroxy group of ribose in stabilizing more complex folding, as been also shown in this thesis.

ABBREVIATIONS

CD = Circular dichroism

Dist-A = Distamycin A

D-loop = Displacement loop

DNA = Deoxyribonucleic acids

DSC = Differential scanning calorimetry

EDTA = Ethylene diamine tetraacetic acid

hTERT = Human telomerase reverse transcriptase

hTR = Human telomerase RNA

ITC = Isothermal titration calorimetry

MD = Molecular Dynamics

mTBA = Modified thrombin binding aptamer

NHE = Nuclease hypersensitivity element

POT1 = Protection of telomere 1

Rap1 = Repressor/activator protein 1

RNA = Ribonucleic acids

TBA = Thrombin binding aptamer

TFO = Triplex forming oligonucleotide

TIN2 = TRF1-interacting nuclear factor 2

T-loop = Telomere loop

TPE = Telomere position effect

TPP1 = Telomerase associated protein 1

TRAP-PCR = Telomeric repeat amplification protocol-polymerase chain reaction

TRF = Telomeric repeat binding factor

UV = Ultra violet

REFERENCES

[1]   Voet, D.; Voet, J. G. *Biochemistry*, 2nd Ed. John Wiley & Sons, Inc. **1995**.
[2]   Watson, J. D.; Crick, F. H. *Nature*, **1953**, *171*, 737-738.
[3]   Felsenfeld, G.; Davies, D. R.; Rich, A. *J. Am. Chem. Soc.*, **1957**, *79*, 2023-2024.
[4]   Lyamichev, V. I.; Mirkin, S. N.; Frank-Kamenetsckii, M. D. *J. Biomol. Struct. Dyn.*, **1985**, *3*, 667-669.
[5]   Hoogsteen, K. *Acta Cryst.*, **1959**, *12*, 822-823.
[6]   Macaya, R.; Wang, E.; Schultze, P.; Sklenar, V.; Feigon, J. *J. Mol. Biol.*, **1992**, *225*, 755-773.
[7]   Howard, F. B.; Miles, H. T.; Liu, K.; Frazier, J.; Raughunathan, G.; Sasisekharan, V. *Biochemistry*, **1992**, *31*, 10671-10677.
[8]   Liu, K.; Sasisekharan, V.; Miles, H. T.; Raughunathan, G. *Biopolymers*, **1996**, *39*, 573-589.
[9]   Arnott, S.; Selsing, E. *J. Mol. Biol.*, **1974**, *88*, 509-521.
[10]   Gellert, M.; Lipsett, M. N.; Davies, D. R. *Proc. Natl. Acad. Sci. USA*, **1962**, *48*, 2013-2018.
[11]   Simonsson, T. *Biol Chem*, **2001**, *382*, 621-628.
[12]   Keniry, M. A. *Biopolymers*, **2001**, *56*, 123-146.
[13]   Parkinson, G. N. In *Quadruplex Nucleic Acids*, S. Neidle; S. Balasubramanian, eds.; RSC Publishing: London, **2006**, pp. 1-30.
[14]   Williamson, J. R. *Annu. Rev. Biophys. Biomol. Struct.*, **1994**, *23*, 703-730.
[15]   Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Nucleic Acids Res.*, **2006**, *34*, 5402-5415.
[16]   Aboul-ela, F.; Murchie, A. I.; Lilley, D. M. *Nature*, **1992**, *360*, 280-282.
[17]   Phillips, K.; Dauter, Z.; Murchie, A. I.; Lilley, D. M.; Luisi, B. *J. Mol. Biol.*, **1997**, *273*, 171-182.
[18]   Strahan, G. D.; Shafer, R. H.; Keniry, M. A. *Nucleic Acids Res*, **1994**, *22*, 5447-5455.
[19]   Smith, F. W.; Lau, F. W.; Feigon, J. *Proc Natl Acad Sci U S A*, **1994**, *91*, 10546-10550.
[20]   Smith, F. W.; Schultze, P.; Feigon, J. *Structure*, **1995**, *3*, 997-1008.
[21]   Esposito, V.; Galeone, A.; Mayol, L.; Oliviero, G.; Virgilio, A.; Randazzo, L. *Nucleosides Nucleotides Nucleic Acids*, **2007**, *26*, 1155-1159.
[22]   Parkinson, G. N.; Lee, M. P.; Neidle, S. *Nature*, **2002**, *417*, 876-880.
[23]   Phan, A. T.; Kuryavyi, V.; Ma, J. B.; Faure, A.; Andreola, M. L.; Patel, D. J. *Proc. Natl. Acad. Sci. USA*, **2005**, *102*, 634-639.
[24]   Smith, F. W.; Feigon, J. *Nature*, **1992**, *356*, 164-168.
[25]   Schultze, P.; Smith, F. W.; Feigon, J. *Structure*, **1994**, *2*, 221-233.
[26]   Schultze, P.; Hud, N. V.; Smith, F. W.; Feigon, J. *Nucleic Acids Res.*, **1999**, *27*, 3018-3028.
[27]   Haider, S.; Parkinson, G. N.; Neidle, S. *J. Mol. Biol.*, **2002**, *320*, 189-200.
[28]   Crnugelj, M.; Sket, P.; Plavec, J. *J. Am. Chem. Soc.*, **2003**, *125*, 7866-7871.
[29]   Crnugelj, M.; Hud, N. V.; Plavec, J. *J. Mol. Biol.*, **2002**, *320*, 911-924.
[30]   Kerwin, S. M. *Curr. Pharm. Des.*, **2000**, *6*, 441-478.
[31]   Wang, Y.; Patel, D. J. *J. Mol. Biol.*, **1995**, *251*, 76-94.
[32]   Kang, C.; Zhang, X.; Ratliff, R.; Moyzis, R.; Rich, A. *Nature*, **1992**, *356*, 126-131.
[33]   Sasisekharan, V.; Zimmerman, S.; Davies, D. R. *J. Mol. Biol.*, **1975**, *92*, 171-174.
[34]   Pinnavaia, T. J.; Marshall, C. L.; Mettler, C. M.; Fisk, C. L.; Miles, H. T.; Becker, E. D. *J. Am. Chem. Soc.*, **1978**, *100*, 3625-3627.
[35]   Simonsson, T.; Sjoback, R. *J. Biol. Chem.*, **1999**, *274*, 17379-17383.
[36]   Laughlan, G.; Murchie, A. I.; Norman, D. G.; Moore, M. H.; Moody, P. C.; Lilley, D. M.; Luisi, B. *Science*, **1994**, *265*, 520-524.
[37]   Chen, F. M. *Biochemistry*, **1992**, *31*, 3769-3776.
[38]   Guschlbauer, W.; Chantot, J. F.; Thiele, D. *J. Biomol. Struct. Dyn.*, **1990**, *8*, 491-511.
[39]   Hardin, C. C.; Watson, T.; Corregan, M.; Bailey, C. *Biochemistry*, **1992**, *31*, 833-841.
[40]   Strahan, G. D.; Keniry, M. A.; Shafer, R. H. *Biophys. J.*, **1998**, *75*, 968-981.

[41] Hud, N. V.; Smith, F. W.; Anet, F. A.; Feigon, J. *Biochemistry*, **1996**, *35*, 15383-15390.

[42] Bouaziz, S.; Kettani, A.; Patel, D. J. *J. Mol. Biol.*, **1998**, *282*, 637.

[43] Kettani, A.; Bouaziz, S.; Gorin, A.; Zhao, A.; Jones, R. A.; Patel, D. J. *J. Mol. Biol.*, **1998**, *282*, 619.

[44] Gill, M. L.; Strobel, S. A.; Loria, J. P. *J. Am. Chem. Soc.*, **2005**, *127*, 16723-16732.

[45] Klobutcher, L. A.; Swanton, M. T.; Donini, P.; Prescott, D. M. *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 3015-3019.

[46] Moyzis, R. K.; Buckingham, J. M.; Cram, L. S.; Dani, M.; Deaven, L. L.; Jones, M. D.; Meyne, J.; Ratliff, R. L.; Wu, J. R. *Proc. Natl. Acad. Sci. USA*, **1988**, *85*, 6622-6626.

[47] Richards, E. J.; Ausubel, F. M. *Cell*, **1988**, *53*, 127-136.

[48] Greider, C. W.; Blackburn, E. H. *Nature*, **1989**, *337*, 331-337.

[49] Kipling, D.; Cooke, H. J. *Nature*, **1990**, *347*, 400-402.

[50] McEachern, M. J.; Hicks, J. B. *Mol. Cell. Biol. *, **1993**, *13*, 551-560.

[51] Prowse, K. R.; Greider, C. W. *Proc. Natl. Acad. Sci. USA*, **1995**, *92*, 4818-4822.

[52] Wang, Y.; Patel, D. J. *Structure*, **1993**, *1*, 263-282.

[53] Wang, Y.; Patel, D. J. *Structure*, **1994**, *2*, 1141-1156.

[54] Mokbel, K.; Parris, C. N.; Ghilchik, M.; Williams, G.; Newbold, R. F. *Am. J. Surg.*, **1999**, *178*, 69-72.

[55] Zhu, J.; Wang, H.; Bishop, J. M.; Blackburn, E. H. *Proc. Natl. Acad. Sci. USA*, **1999**, *96*, 3723-3728.

[56] Mokbel, K. *Curr. Med. Res. Opin.*, **2003**, *19*, 470-472.

[57] Hahn, W. C.; Counter, C. M.; Lundberg, A. S.; Beijersbergen, R. L.; Brooks, M. W.; Weinberg, R. A. *Nature*, **1999**, *400*, 464-468.

[58] Sun, D.; Thompson, B.; Cathers, B. E.; Salazar, M.; Kerwin, S. M.; Trent, J. O.; Jenkins, T. C.; Neidle, S.; Hurley, L. H. *J. Med. Chem.*, **1997**, *40*, 2113-2116.

[59] Bryan, T. M.; Cech, T. R. *Curr. Opin. Cell Biol.*, **1999**, *11*, 318-324.

[60] de Lange, T.; Jacks, T. *Cell*, **1999**, *98*, 273-275.

[61] Perry, P. J.; Read, M. A.; Davies, R. T.; Gowan, S. M.; Reszka, A. P.; Wood, A. A.; Kelland, L. R.; Neidle, S. *J. Med. Chem.*, **1999**, *42*, 2679-2684.

[62] Gowan, S. M.; Heald, R.; Stevens, M. F.; Kelland, L. R. *Mol. Pharmacol.*, **2001**, *60*, 981-988.

[63] Mergny, J.-L.; Lacroix, L.; Teulade-Fichou, M.-P.; Hounsou, C.; Guittat, L.; Hoarau, M.; Arimondo, P.; Vigneron, J.-P.; Lehn, J.-M.; Riou, J.-F.; Garestier, T.; Helene, C. *Proc. Natl. Acad. Sci. USA*, **2001**, *98*, 3062-3067.

[64] Mergny, J. L.; Riou, J. F.; Mailliet, P.; Teulade-Fichou, M. P.; Gilson, E. *Nucleic Acids Res.*, **2002**, *30*, 839-865.

[65] Rezler, E. M.; Bearss, D. J.; Hurley, L. H. *Curr. Opin. Pharmacol.*, **2002**, *2*, 415-423.

[66] Riou, J. F.; Guittat, L.; Mailliet, P.; Laoui, A.; Renou, E.; Petitgenet, O.; Megnin-Chanet, F.; Helene, C.; Mergny, J. L. *Proc. Natl. Acad. Sci. USA*, **2002**, *99*, 2672-2677.

[67] Rezler, E. M.; Bearss, D. J.; Hurley, L. H. *Annu. Rev. Pharmacol. Toxicol.*, **2003**, *43*, 359-379.

[68] Schaffitzel, C.; Berger, I.; Postberg, J.; Hanes, J.; Lipps, H. J.; Pluckthun, A. *Proc. Natl. Acad. Sci. USA*, **2001**, *98*, 8572-8577.

[69] Chang, C. C.; Chu, J. F.; Kao, F. J.; Chiu, Y. C.; Lou, P. J.; Chen, H. C.; Chang, T. C. *Anal Chem*, **2006**, *78*, 2810-2815.

[70] Simonsson, T.; Pecinka, P.; Kubista, M. *Nucleic Acids Res.*, **1998**, *26*, 1167-1172.

[71] Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. *Proc. Natl. Acad. Sci. USA*, **2002**, *99*, 11593-11598.

[72] Phan, A. T.; Modi, Y. S.; Patel, D. J. *J. Am. Chem. Soc.*, **2004**, *126*, 8710-8716.

[73] Ambrus, A.; Chen, D.; Dai, J.; Jones, R. A.; Yang, D. *Biochemistry*, **2005**, *44*, 2048-2058.

[74] Rankin, S.; Reszka, A. P.; Huppert, J.; Zloh, M.; Parkinson, G. N.; Todd, A. K.; Ladame, S.; Balasubramanian, S.; Neidle, S. *J. Am. Chem. Soc.*, **2005**, *127*, 10584-10589.

[75] Fernando, H.; Reszka, A. P.; Huppert, J.; Ladame, S.; Rankin, S.; Venkitaraman, A. R.; Neidle, S.; Balasubramanian, S. *Biochemistry*, **2006**, *45*, 7854-7860.

[76] De Armond, R.; Wood, S.; Sun, D.; Hurley, L. H.; Ebbinghaus, S. W. *Biochemistry*, **2005**, *44*, 16341-16350.

[77] Sun, D.; Guo, K.; Rusche, J. J.; Hurley, L. H. *Nucleic Acids Res.*, **2005**, *33*, 6070-6080.

[78] Dai, J.; Chen, D.; Jones, R. A.; Hurley, L. H.; Yang, D. *Nucleic Acids Res.*, **2006**, *34*, 5133-5144.

[79] Dai, J.; Dexheimer, T. S.; Chen, D.; Carver, M.; Ambrus, A.; Jones, R. A.; Yang, D. *J. Am. Chem. Soc.*, **2006**, *128*, 1096-1098.

[80] Cogoi, S.; Xodo, L. E. *Nucleic Acids Res.*, **2006**, *34*, 2536-2549.

[81] Murchie, A. I.; Lilley, D. M. *Nucleic Acids Res.*, **1992**, *20*, 49-53.

[82] Xu, Y.; Sugiyama, H. *Nucleic Acids Res.*, **2006**, *34*, 949-954.

[83] Sun, D.; Pourpak, A.; Beetz, K.; Hurley, L. H. *Clin. Cancer Res.*, **2003**, *9*, A218.

[84] Spencer, C. A.; Groudine, M. *Adv. Cancer Res.*, **1991**, *56*, 1-48.

[85] Marcu, K. B.; Bossone, S. A.; Patel, A. J. *Annu. Rev. Biochem.*, **1992**, *61*, 809-860.

[86] Facchini, L. M.; Penn, L. Z. *FASEB J.*, **1998**, *12*, 633-651.

[87] Pelengaris, S.; Rudolph, B.; Littlewood, T. *Curr. Opin. Genet. Dev.*, **2000**, *10*, 100-105.

[88] Wang, J.; Xie, L. Y.; Allan, S.; Beach, D.; Hannon, G. J. *Genes Dev.*, **1998**, *12*, 1769-1774.

[89] Bossone, S. A.; Asselin, C.; Patel, A. J.; Marcu, K. B. *Proc. Natl. Acad. Sci. USA*, **1992**, *89*, 7452-7456.

[90] Cooney, M.; Czernuszewicz, G.; Postel, E. H.; Flint, S. J.; Hogan, M. E. *Science*, **1988**, *241*, 456-459.

[91] Postel, E. H.; Berberich, S. J.; Rooney, J. W.; Kaetzel, D. M. *J. Bioenerg. Biomembr.*, **2000**, *32*, 277-284.

[92] Sakatsume, O.; Tsutsui, H.; Wang, Y.; Gao, H.; Tang, X.; Yamauchi, T.; Murata, T.; Itakura, K.; Yokoyama, K. *J. Biol. Chem.*, **1996**, *271*, 31322-31333.

[93] Tomonaga, T.; Levens, D. *Proc. Natl. Acad. Sci. USA*, **1996**, *93*, 5830-5835.

[94] Collins, I.; Weber, A.; Levens, D. *Mol. Cell. Biol.*, **2001**, *21*, 8437-8451.

[95] Seenisamy, J.; Bashyam, S.; Gokhale, V.; Vankayalapati, H.; Sun, D.; Siddiqui-Jain, A.; Streiner, N.; Shin-Ya, K.; White, E.; Wilson, W. D.; Hurley, L. H. *J. Am. Chem. Soc.*, **2005**, *127*, 2944-2959.

[96] Lin, Y.; Padmapriya, A.; Morden, K. M.; Jayasena, S. D. *Proc. Natl. Acad. Sci. USA*, **1995**, *92*, 11044-11048.

[97] Lin, Y.; Jayasena, S. D. *J. Mol. Biol.*, **1997**, *271*, 100-111.

[98] Wen, J. D.; Gray, D. M. *Biochemistry*, **2002**, *41*, 11438-11448.

[99] Griffin, L. C.; Tidmarsh, G. F.; Bock, L. C.; Toole, J. J.; Leung, L. L. *Blood*, **1993**, *81*, 3271-3276.

[100] Li, W. X.; Kaplan, A. V.; Grant, G. W.; Toole, J. J.; Leung, L. L. *Blood*, **1994**, *83*, 677-682.

[101] Jing, N.; Hogan, M. E. *J. Biol. Chem.*, **1998**, *273*, 34992-34999.

[102] Qi, H.; Lin, C. P.; Fu, X.; Wood, L. M.; Liu, A. A.; Tsai, Y. C.; Chen, Y.; Barbieri, C. M.; Pilch, D. S.; Liu, L. F. *Cancer Res.*, **2006**, *66*, 11808-11816.

[103] Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.*, **1990**, *112*, 1393-1399.

[104] Randazzo, A.; Galeone, A.; Mayol, L. *Chem. Commun.*, **2001**, *11*, 1030-1031.

[105] Cocco, M. J.; Hanakahi, L. A.; Huber, M. D.; Maizels, N. *Nucl. Acids Res.*, **2003**, *31*, 2944-2951.

[106] Martino, L.; Virno, A.; Randazzo, A.; Virgilio, A.; Esposito, V.; Giancola, C.; Bucci, M.; Cirino, G.; Mayol, L. *Nucleic Acids Res.*, **2006**, *34*, 6653-6662.

[107] Uesugi, S.; Liu, H.; Kugimiya, A.; Matsugami, A.; Katahira, M. *Nucleic Acids Res. Suppl.*, **2003**, 51-52.

[108] Liu, H.; Matsugami, A.; Katahira, M.; Uesugi, S. *J. Mol. Biol.*, **2002**, *322*, 955-970.

[109] Doyle, M. L. *Curr Opin Biotechnol*, **1997**, *8*, 31-35.

[110] Ababou, A.; Ladbury, J. E. *J Mol Recognit*, **2006**, *19*, 79-89.

[111] Pierce, M. M.; Raman, C. S.; Nall, B. T. *Methods*, **1999**, *19*, 213-221.

[112] Haq, I.; Ladbury, J. *J Mol Recognit*, **2000**, *13*, 188-197.

[113] Leavitt, S.; Freire, E. *Curr. Opin. Struct. Biol.*, **2001**, *11*, 560-566.

[114] O'Brien, R.; Ladbury, J. E.; Chowdry, B. Z. In *Protein-Ligand Interactions: hydrodynamics and calorimetry*, S. E. Harding; B. Z. Chowdry, eds.; Oxford University Press: Oxford, **2001**, pp. 263-286.

[115] Wiseman, T.; Williston, S.; Brants, J. F.; Lin, L. N. *Anal. Biochem.*, **1989**, *179*, 131-137.

[116] Giancola, C. *J. Therm. Anal. Cal.*, **2007**, *91*, online.

[117] Barone, G.; Del Vecchio, P.; Fessas, D.; Giancola, C.; Graziano, G. *J. Thermal. Anal.*, **1993**, *38*, 2779.

[118] Bruylants, G.; Wouters, J.; Michaux, C. *Curr. Med. Chem.*, **2005**, *12*, 2011-2020.

[119] Betts, M. J.; Sternberg, M. J. *Protein Eng*, **1999**, *12*, 271-283.

[120] Totrov, M.; Abagyan, R. *Nat Struct Biol*, **1994**, *1*, 259-263.

[121] Strynadka, N. C.; Eisenstein, M.; Katchalski-Katzir, E.; Shoichet, B. K.; Kuntz, I. D.; Abagyan, R.; Totrov, M.; Janin, J.; Cherfils, J.; Zimmerman, F.; Olson, A.; Duncan, B.; Rao, M.; Jackson, R.; Sternberg, M.; James, M. N. *Nat Struct Biol*, **1996**, *3*, 233-239.

[122] Chen, R.; Li, L.; Weng, Z. *Proteins*, **2003**, *52*, 80-87.

[123] Chen, R.; Weng, Z. *Proteins*, **2002**, *47*, 281-294.

[124] Zhang, C.; Vasmatzis, G.; Cornette, J. L.; DeLisi, C. *J Mol Biol*, **1997**, *267*, 707-726.

[125] Zhang, C.; Cornette, J. L.; Delisi, C. *Protein Sci*, **1997**, *6*, 1057-1064.

[126] Vasmatzis, G.; Zhang, C.; Cornette, J. L.; DeLisi, C. *Mol Immunol*, **1996**, *33*, 1231-1239.

[127] Zhang, C.; Chen, J.; DeLisi, C. *Proteins*, **1999**, *34*, 255-267.

[128] Gabb, H.; Jackson, R.; Sternberg, M. *Journal of Molecular Biology*, **1997**, *272*, 106-120.

[129] Aloy, P.; Moont, G.; Gabb, H. A.; Querol, E.; Aviles, F. X.; Sternberg, M. J. *Proteins*, **1998**, *33*, 535-549.

[130] Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. *Proteins*, **2002**, *47*, 409-443.

[131] De Luca, L.; Pedretti, A.; Vistoli, G.; Barreca, M. L.; Villa, L.; Monforte, P.; Chimirri, A. *Biochem Biophys Res Commun*, **2003**, *310*, 1083-1088.

[132] Adesokan, A. A.; Roberts, V. A.; Lee, K. W.; Lins, R. D.; Briggs, J. M. *J Med Chem*, **2004**, *47*, 821-828.

[133] Roberts, V. A.; Case, D. A.; Tsui, V. *Proteins*, **2004**, *57*, 172-187.

[134] Zhu, H. M.; Chen, W. Z.; Wang, C. X. *Bioorg Med Chem Lett*, **2005**, *15*, 475-477.

[135] Van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glattli, A.; Hunenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F. A.; Yu, H. B. *Angew. Chem. Int. Ed.*, **2006**, *45*, 4064-4092.

[136] Griffith, J. D.; Comeau, L.; Rosenfield, S.; Stansel, R. M.; Bianchi, A.; Moss, H.; de Lange, T. *Cell*, **1999**, *97*, 503-514.

[137] Blackburn, E. H. *Nature*, **2000**, *408*, 53-56.

[138] Blackburn, E. H. *Cell*, **2001**, *106*, 661-673.

[139] de Lange, T. *Oncogene*, **2002**, *21*, 532-540.

[140] Rodier, F.; Kim, S. H.; Nijjar, T.; Yaswen, P.; Campisi, J. *Int J Biochem Cell Biol*, **2005**, *37*, 977-990.

[141] Pendino, F.; Tarkanyi, I.; Dudognon, C.; Hillion, J.; Lanotte, M.; Aradi, J.; Ségal-Bendirdjian, E. *Curr. Cancer Drug Targets*, **2006**, *6*, 147-180.

[142] Allshire, R. C.; Dempster, M.; Hastie, N. D. *Nucleic Acids Res.*, **1989**, *17*, 4611-4627.

[143] de Lange, T.; Shiue, L.; Myers, R. M.; Cox, D. R.; Naylor, S. L.; Killery, A. M.; Varmus, H. E. *Mol Cell Biol*, **1990**, *10*, 518-527.

[144] de Lange, T. *Genes Dev*, **2005**, *19*, 2100-2110.

[145] Zhong, Z.; Shiue, L.; Kaplan, S.; de Lange, T. *Mol. Cell. Biol.*, **1992**, *12*, 4834-4843.

[146] Chong, L.; van Steensel, B.; Broccoli, D.; Erdjument-Bromage, H.; Hanish, J.; Tempst, P.; de Lange, T. *Science*, **1995**, *270*, 1663-1667.

[147] Bilaud, T.; Brun, C.; Ancelin, K.; Koering, C. E.; Laroche, T.; Gilson, E. *Nat. Genet.*, **1997**, *17*, 236-239.

[148] Broccoli, D.; Smogorzewska, A.; Chong, L.; de Lange, T. *Nat. Genet.*, **1997**, *17*, 231-235.

[149] Kim, S. H.; Kaminker, P.; Campisi, J. *Nat. Genet.*, **1999**, *23*, 405-412.

[150] Baumann, P.; Cech, T. R. *Science*, **2001**, *292*, 1171-1175.

[151] Colgin, L. M.; Baran, K.; Baumann, P.; Cech, T. R.; Reddel, R. R. *Curr Biol*, **2003**, *13*, 942-946.

[152] Kelleher, C.; Kurth, I.; Lingner, J. *Mol Cell Biol*, **2005**, *25*, 808-818.

[153] O'Connor, M. S.; Safari, A.; Xin, H.; Liu, D.; Songyang, Z. *Proc Natl Acad Sci U S A*, **2006**, *103*, 11874-11879.

[154] Blasco, M. A. *Nat. Rev. Genet.*, **2005**, *6*, 611-622.

[155] Songyang, Z.; Liu, D. *Crit Rev Eukaryot Gene Expr*, **2006**, *16*, 103-118.

[156] Harley, C. B.; Futcher, A. B.; Greider, C. W. *Nature*, **1990**, *345*, 458-460.

[157] Allsopp, R. C.; Harley, C. B. *Exp Cell Res*, **1995**, *219*, 130-136.

[158] Kim, N. W. *Eur J Cancer*, **1997**, *33*, 781-786.

[159] Weinrich, S. L.; Pruzan, R.; Ma, L.; Ouellette, M.; Tesmer, V. M.; Holt, S. E.; Bodnar, A. G.; Lichtsteiner, S.; Kim, N. W.; Trager, J. B.; Taylor, R. D.; Carlos, R.; Andrews, W. H.; Wright, W. E.; Shay, J. W.; Harley, C. B.; Morin, G. B. *Nat Genet*, **1997**, *17*, 498-502.

[160] Shay, J. W.; Wright, W. E. *Cancer Cell*, **2002**, *2*, 257-265.

[161] Feng, J.; Funk, W. D.; Wang, S. S.; Weinrich, S. L.; Avilion, A. A.; Chiu, C. P.; Adams, R. R.; Chang, E.; Allsopp, R. C.; Yu, J.; Le, S. Y.; West, M. D.; Harley, C. B.; Andrews, W. H.; Greider, C. W.; Villeponteau, B. *Science*, **1995**, *269*, 1236-1241.

[162] Nakamura, T. M.; Morin, G. B.; Chapman, K. B.; Weinrich, S. L.; Andrews, W. H.; Lingner, J.; Harley, C. B.; Cech, T. R. *Science*, **1997**, *277*, 955-959.

[163] Beattie, T. L.; Zhou, W.; Robinson, M. O.; Harrington, L. *Mol Cell Biol*, **2001**, *21*, 6151-6160.

[164] Blackburn, E. H. *FEBS Lett.*, **2005**, *579*, 859-862.

[165] Cesare, A. J.; Griffith, J. D. *Mol. Cell. Biol.*, **2004**, *24*, 9948-9957.

[166] Hahn, W. C. *Curr. Mol. Med.*, **2005**, *5*, 227-231.

[167] LaBranche, H.; Dupuis, S.; Ben-David, Y.; Bani, M. R.; Wellinger, R. J.; Chabot, B. *Nat. Genet.*, **1998**, *19*, 199-202.

[168] Nosek, J.; Rycovska, A.; Makhov, A. M.; Griffith, J. D.; Tomaska, L. *J. Biol. Chem.*, **2005**, *280*, 10840-10845.

[169] Beattie, T. L.; Zhou, W.; Robinson, M. O.; Harrington, L. *Curr Biol*, **1998**, *8*, 177-180.

[170] Nugent, C. I.; Lundblad, V. *Genes Dev*, **1998**, *12*, 1073-1085.

[171] Kim, N. W.; Piatyszek, M. A.; Prowse, K. R.; Harley, C. B.; West, M. D.; Ho, P. L.; Coviello, G. M.; Wright, W. E.; Weinrich, S. L.; Shay, J. W. *Science*, **1994**, *266*, 2011-2015.

[172] Shay, J. W.; Gazdar, A. F. *J Clin Pathol*, **1997**, *50*, 106-109.

[173] Sharma, S.; Raymond, E.; Soda, H.; Sun, D.; Hilsenbeck, S. G.; Sharma, A.; Izbicka, E.; Windle, B.; Von Hoff, D. D. *Ann Oncol*, **1997**, *8*, 1063-1074.

[174] Mergny, J. L.; Helene, C. *Nat. Med.*, **1998**, *4*, 1366-1367.

[175] Neidle, S.; Kelland, L. R. *Anticancer Drug Des*, **1999**, *14*, 341-347.

[176] Kelland, L. R. *Lancet Oncol*, **2001**, *2*, 95-102.

[177] Alberti, P.; Lacroix, L.; Guittat, L.; Helene, C.; Mergny, J. L. *Mini Rev. Med. Chem.*, **2003**, *3*, 23-36.

[178] Gellert, G. C.; Jackson, S. R.; Dikmen, Z. G.; Wright, W. E.; Shay, J. W. *Drug Discovery Today: Disease Mechanism* **2005**, *2*, 159-164.

[179] McKenzie, K. E.; Umbricht, C. B.; Sukumar, S. *Mol Med Today*, **1999**, *5*, 114-122.

[180] Read, M. A.; Wood, A. A.; Harrison, J. R.; Gowan, S. M.; Kelland, L. R.; Dosanjh, H. S.; Neidle, S. *J Med Chem*, **1999**, *42*, 4538-4546.

[181] Hodes, R. *Proc. Natl. Acad. Sci. USA*, **2001**, *98*, 7649-7651.

[182] Kim, M. M.; Rivera, M. A.; Botchkina, I. L.; Shalaby, R.; Thor, A. D.; Blackburn, E. H. *Proc Natl Acad Sci U S A*, **2001**, *98*, 7982-7987.

[183] White, L. K.; Wright, W. E.; Shay, J. W. *Trends Biotechnol*, **2001**, *19*, 114-120.

[184] Gowan, S. M.; Harrison, J. R.; Patterson, L.; Valenti, M.; Read, M. A.; Neidle, S.; Kelland, L. R. *Mol. Pharmacol.*, **2002**, *61*, 1154-1162.

[185] Helder, M. N.; Wisman, G. B.; van der Zee, G. J. *Cancer Invest*, **2002**, *20*, 82-101.

[186] Neidle, S.; Parkinson, G. *Nat Rev Drug Discov*, **2002**, *1*, 383-393.

[187] Saretzki, G. *Cancer Letters*, **2003**, *194*, 209-219.

[188] Dikmen, Z. G.; Gellert, G. C.; Jackson, S.; Gryaznov, S.; Tressler, R.; Dogan, P.; Wright, W. E.; Shay, J. W. *Cancer Res.*, **2005**, *65*, 7866-7873.

[189] Gilley, D.; Tanaka, H.; Herbert, B.-S. *Int. J. Biochem. Cell Biol.*, **2005**, *37*, 1000-1013.

[190] Shay, J. *Cancer Res*, **2005**, *65*, 3513-3517.

[191] Shay, J.; Wright, W. *Nat. Rev. Drug Discov.*, **2006**, *5*, 577-584.

[192] Brunsvig, P. F.; Aamdal, S.; Gjertsen, M. K.; Kvalheim, G.; Markowski-Grimsrud, C. J.; Sve, I.; Dyrhaug, M.; Trachsel, S.; Moller, M.; Eriksen, J. A.; Gaudernack, G. *Cancer Immunol. Immunother.*, **2006**, *55*, 1553-1564.

[193] Vonderheide, R. H. *Oncogene*, **2002**, *21*, 674-679.

[194] Vonderheide, R. H.; Domchek, S. M.; Schultze, J. L.; George, D. J.; Hoar, K. M.; Chen, D. Y.; Stephans, K. F.; Masutomi, K.; Loda, M.; Xia, Z.; Anderson, K. S.; Hahn, W. C.; Nadler, L. M. *Clin. Cancer Res.*, **2004**, *10*, 828-839.

[195] Corey, D. R. *Chem. Res. Toxicol.*, **2000**, *13*, 957-960.

[196] Djojosubroto, M. W.; Chin, A. C.; Go, N.; Schaetzlein, S.; Manns, M. P.; Gryaznov, S.; Harley, C. B.; Rudolph, K. L. *Hepatology*, **2005**, *42*, 1127-1136.

[197] Gellert, G. C.; Dikmen, Z. G.; Wright, W. E.; Gryaznov, S.; Shay, J. W. *Breast Cancer Res. Treat.*, **2006**, *96*, 73-81.

[198] Lin, T.; Huang, X.; Gu, J.; Zhang, L.; Roth, J.; Xiong, M.; Curley, S.; Yu, Y.; Hunt, K.; Fang. *Oncogene*, **2002**, *21*, 8020-8028.

[199] Bilsland, A. E.; Anderson, C. J.; Fletcher-Monaghan, A. J.; McGregor, F.; Evans, T. R.; Ganly, I.; Knox, R. J.; Plumb, J. A.; Keith, W. N. *Oncogene*, **2003**, *22*, 370-380.

[200] Wirth, T.; Zender, L.; Schulte, B.; Mundt, B.; Plentz, R.; Rudolph, K. L.; Manns, M.; Kubicka, S.; Kuhnel, F. *Cancer Res.*, **2003**, *63*, 3181-3188.

[201] Kawashima, T.; Kagawa, S.; Kobayashi, N.; Shirakiya, Y.; Umeoka, T.; Teraishi, F.; Taki, M.; Kyo, S.; Tanaka, N.; Fujiwara, T. *Clin. Cancer Res.*, **2004**, *10*, 285-292.

[202] Yeo, M.; Rha, S. Y.; Jeung, H. C.; Hu, S. X.; Yang, S. H.; Kim, Y. S.; An, S. W.; Chung, H. C. *Int. J. Cancer*, **2005**, *114*, 484-489.

[203] Nosrati, M.; Li, S.; Bagheri, S.; Ginzinger, D.; Blackburn, E.; Debs, R.; Kashani-Sabet, M. *Clin. Cancer Res.*, **2004**, *10*, 4983-4990.

[204] Zahler, A. M.; Williamson, J. R.; Cech, T. R.; Prescott, D. M. *Nature*, **1991**, *350*, 718-720.

[205] Li, J.; Correia, J. J.; Wang, L.; Trent, J. O.; Chaires, J. B. *Nucleic Acids Res.*, **2005**, *33*, 4649-4659.

[206] Ying, L.; Green, J. J.; Li, H.; Klenerman, D.; Balasubramanian, S. *Proc. Natl. Acad. Sci. USA*, **2003**, *100*, 14629-14634.

[207] Lee, J. Y.; Okumus, B.; Kim, D. S.; Ha, T. *Proc. Natl. Acad. Sci. USA*, **2005**, *102*, 18938-18943.

[208] Ourliac-Garnier, I.; Elizondo-Riojas, M. A.; Redon, S.; Farrell, N. P.; Bombard, S. *Biochemistry*, **2005**, *44*, 10620-10634.

[209] Rujan, I. N.; Meleney, J. C.; Bolton, P. H. *Nucleic Acids Res.*, **2005**, *33*, 2022-2031.

[210] Vorlickova, M.; Chladkova, J.; Kejnovska, I.; Fialova, M.; Kypr, J. *Nucleic Acids Res.*, **2005**, *33*, 5851-5860.

[211] Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D. *Nucleic Acids Res.*, **2006**, *34*, 2723-2735.

[212] Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J. *J. Am. Chem. Soc.*, **2006**, *128*, 9963-9970.

[213] Phan, A. T.; Luu, K. N.; Patel, D. J. *Nucleic Acids Res.*, **2006**, *34*, 5715-5719.

[214] Xu, Y.; Noguchi, Y.; Sugiyama, H. *Bioorg. Med. Chem.*, **2006**, *14*, 5584-5591.

[215] Dai, J.; Punchihewa, C.; Ambrus, A.; Chen, D.; Jones, R. A.; Yang, D. *Nucleic Acids Res.*, **2007**, *35*, 2440-2450.

[216] Balagurumoorthy, P.; Brahmachari, S. K.; Mohanty, D.; Bansal, M.; Sasisekharan, V. *Nucl. Acids Res.*, **1992**, *20*, 4061-4067.

[217] Paramasivan, S.; Rujan, I.; Bolton, P. H. *Methods*, **2007**, *43*, 324–331.

[218] Aboul-ela, F.; Murchie, A. I.; Norman, D. G.; Lilley, D. M. *J. Mol. Biol.*, **1994**, *243*, 458-471.

[219] Zaffaroni, N.; Lualdi, S.; Villa, R.; Bellarosa, D.; Cermele, C.; Felicetti, P.; Rossi, C.; Orlandi, L.; Daidone, M. G. *Eur. J. Cancer*, **2002**, *38*, 1792-1801.

[220] Moore, M. J.; Cuenca, F.; Searcey, M.; Neidle, S. *Org. Biomol. Chem.*, **2006**, *4*, 3479-3488.

[221] Kim, M. Y.; Gleason-Guzman, M.; Izbicka, E.; Nishioka, D.; Hurley, L. H. *Cancer Res.*, **2003**, *63*, 3247-3256.

[222] Izbicka, E.; Wheelhouse, R. T.; Raymond, E.; Davidson, K. K.; Lawrence, R. A.; Sun, D.; Windle, B. E.; Hurley, L. H.; Von Hoff, D. D. *Cancer Res.*, **1999**, *59*, 639-644.

[223] Ren, J.; Chaires, J. B. *Biochemistry* **1999**, *38*, 16067-16075.

[224] Wang, P.; Ren, L.; He, H.; Liang, F.; Zhou, X.; Tan, Z. *Chembiochem*, **2006**, *7*, 1155-1159.

[225] Wei, C.; Jia, G.; Yuan, J.; Feng, Z.; Li, C. *Biochemistry*, **2006**, *45*, 6681-6691.

[226] Mita, H.; Ohyama, T.; Tanaka, Y.; Yamamoto, Y. *Biochemistry*, **2006**, *45*, 6765-6772.

[227] Han, H.; Langley, D. R.; Rangan, A.; Hurley, L. H. *J. Am. Chem. Soc.*, **2001**, *123*, 8902-8913.

[228] Parkinson, G. N.; Ghosh, R.; Neidle, S. *Biochemistry*, **2007**, *46*, 2390-2397.

[229] Cantor, C. R.; Warshow, M. M.; Shapiro, H. *Biopolymers*, **1970**, *9*, 1059-1077.

[230] Animati, F.; Arcamone, F. M.; Conte, M. R.; Felicetti, P.; Galeone, A.; Lombardi, P.; Mayol, L.; Paloma, L. G.; Rossi, C. *J. Med. Chem.*, **1995**, *38*, 1140-1149.

[231] Erra, E.; Petraccone, L.; Esposito, V.; Randazzo, A.; Mayol, L.; Ladbury, J.; Barone, G.; Giancola, C. *Nucleosides Nucleotides Nucleic Acids*, **2005**, *24*, 753-756.

[232] Haq, I.; Trent, J. O.; Chowdhry, B. Z.; Jenkins, T. C. *J. Am. Chem. Soc.*, **1999**, *121*, 1768-1779.

[233] Haider, S. M.; Parkinson, G. N.; Neidle, S. *J. Mol. Biol.*, **2003**, *326*, 117-125.

[234] Gavathiotis, E.; Heald, R. A.; Stevens, M. F.; Searle, M. S. *J. Mol. Biol.*, **2003**, *334*, 25-36.

[235] White, E. W.; Tanious, F.; Ismail, M. A.; Reszka, A. P.; Neidle, S.; Boykin, D. W.; Wilson, W. D. *Biophys. Chem.*, **2007**, *126*, 140-153.

[236] Martino, L.; Virno, A.; Pagano, B.; Virgilio, A.; Di Micco, S.; Galeone, A.; Giancola, C.; Bifulco, G.; Mayol, L.; Randazzo, A. *J. Am. Chem. Soc.*, **2007**, *129*, 16048-16056.

[237] Chaires, J. B. *Arch. Biochem. Biophys.*, **2006**, *453*, 26-31.

[238] Chaires, J. B. *Biopolymers*, **1997**, *44*, 201-215.

[239] Haq, I. *Arch. Biochem. Biophys.*, **2002**, *403*, 1-15.

[240] Caceres, C.; Wright, G.; Gouyette, C.; Parkinson, G.; Subirana, J. A. *Nucleic Acids Res.*, **2004**, *32*, 1097-1102.

[241] Creze, C.; Rinaldi, B.; Haser, R.; Bouvet, P.; Gouet, P. *Acta Crystallogr. D Biol. Crystallogr.*, **2007**, *D63*, 682-688.

[242] Chowdhury, S.; Bansal, M. *J. Biomol. Struct. Dyn.*, **2001**, *18*, 647-669.

[243] Chowdhury, S.; Bansal, M. *J. Phys. Chem. B*, **2001**, *105*, 7572-7578.

[244] Davie, E. W.; Fujikawa, K.; Kisiel, W. *Biochemistry*, **1991**, *30*, 10363-10370.

[245] Huntington, J. A. *J. Thromb. Haemost.*, **2005**, *3*, 1861-1872.

[246] Stubbs, M. T.; Bode, W. *Thromb. Res.*, **1993**, *69*, 1-58.

[247] Degen, S. J. F.; Sun, W. Y. *Crit. Rev. Eukaryot. Gene Expr.*, **1998**, *8*, 203-224.

[248] Bock, L. C.; Griffin, L. C.; Latham, J. A.; Vermaas, E. H.; Toole, J. J. *Nature*, **1992**, *355*, 564-566.

[249] Macaya, R. F.; Schultze, P.; Smith, F. W.; Roe, J. A.; Feigon, J. *Proc. Natl. Acad. Sci. USA*, **1993**, *90*, 3745-3749.

[250] Wang, K. Y.; McCurdy, S.; Shea, R. G.; Swaminathan, S.; Bolton, P. H. *Biochemistry*, **1993**, *32*, 1899-1904.

[251] Wu, Q.; Tsiang, M.; Sadler, J. E. *J. Biol. Chem.*, **1992**, *267*, 24408-24412.

[252] Paborsky, L. R.; McCurdy, S. N.; Griffin, L. C.; Toole, J. J.; Leung, L. L. *J. Biol. Chem.*, **1993**, *268*, 20808-20811.

[253] Tsiang, M.; Jain, A. K.; Dunn, K. E.; Rojas, M. E.; Leung, L. L.; Gibbs, C. S. *J. Biol. Chem.*, **1995**, *270*, 16854-16863.

[254] Tsiang, M.; Gibbs, C. S.; Griffin, L. C.; Dunn, K. E.; Leung, L. L. *J. Biol. Chem.*, **1995**, *270*, 19370-19376.

[255] Padmanabhan, K.; Padmanabhan, K. P.; Ferrara, J. D.; Sadler, J. E.; Tulinsky, A. *J. Biol. Chem.*, **1993**, *268*, 17651-17654.

[256] Padmanabhan, K.; Tulinsky, A. *Acta Crystallogr. D Biol. Crystallogr.*, **1996**, *52*, 272-282.

[257] Kelly, J. A.; Feigon, J.; Yeates, T. O. *J. Mol. Biol.*, **1996**, *256*, 417-422.

[258] Wang, K. Y.; Krawczyk, S. H.; Bischofberger, N.; Swaminathan, S.; Bolton, P. H. *Biochemistry*, **1993**, *32*, 11285-11292.

[259] Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.*, **1995**, *91*, 43-56.

[260] Cheatham, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.*, **1999**, *16*, 845-862.

[261] Pearlman, D.; Case, D.; Caldwell, J.; Ross, W.; Cheatham, I.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.*, **1995**, *91*, 1-41.

[262] Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.*, **1983**, *79*, 926-935.

[263] Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. *J. Comput. Chem.*, **1997**, *18*, 1463-1472.

[264] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.*, **1984**, *81*, 3684-3690.

[265] Darden, T.; Perera, L.; Li, L.; Pedersen, L. *Structure*, **1999**, *7*, 55-60.

[266] Fanelli, F.; Ferrari, S. *J. Struct. Biol.*, **2006**, *153*, 278-283.

[267] Olsen, C. M.; Gmeiner, W. H.; Marky, L. A. *J. Phys. Chem. B*, **2006**, *110*, 6962-6969.

[268] Smirnov, I.; Shafer, R. H. *Biochemistry*, **2000**, *39*, 1462-1468.

[269] Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.*, **1988**, *6*, 63-91.

[270] Shafer, R. H.; Smirnov, I. *Biopolymers*, **2001**, *56*, 209-227.

[271] Schaeffer, C.; Bardoni, B.; Mandel, J. L.; Ehresmann, B.; Ehresmann, C.; Moine, H. *EMBO J*, **2001**, *20*, 4803-4813.

[272] Kumari, S.; Bugaut, A.; Huppert, J. L.; Balasubramanian, S. *Nat Chem Biol*, **2007**, *3*, 218-221.

[273] Wieland, M.; Hartig, J. S. *Chem Biol*, **2007**, *14*, 757-763.

[274] Patel, P. K.; Hosur, R. V. *Nucleic Acids Res.*, **1999**, *27*, 2457-2464.

[275] Patel, P. K.; Koti, A. S.; Hosur, R. V. *Nucleic Acids Res.*, **1999**, *27*, 3836-3843.

[276] Zhang, N.; Gorin, A.; Majumdar, A.; Kettani, A.; Chernichenko, N.; Skripin, E.; Patel, D. J. *J. Mol. Biol.*, **2001**, *312*, 1073-1088.

[277] Heller, M.; Flemington, E.; Kieff, E.; Deininger, P. *Mol. Cell. Biol.*, **1985**, *5*, 457-465.

[278] Koch, K. S.; Gleiberman, A. S.; Aoki, T.; Leffert, H. L.; Feren, A.; Jones, A. L.; Fodor, E. J. *Nucleic Acids Res.*, **1995**, *23*, 1098-1112.

[279] Skripkin, E.; Paillart, J. C.; Marquet, R.; Ehresmann, B.; Ehresmann, C. *Proc. Natl. Acad. Sci. USA*, **1994**, *91*, 4945-4949.

[280] Auffinger, P.; Westhof, E. *J. Mol. Biol.*, **2000**, *300*, 1113-1131.

[281] Spackovà, N.; Berger, I.; Sponer, J. *J. Am. Chem. Soc.*, **2001**, *123*, 3295-3307.

[282] Giudice, E.; Lavery, R. *Acc. Chem. Res.*, **2002**, *35*, 350-357.

[283] Cheatham III, T. E.; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. *J. Am. Chem. Soc.*, **1995**, *117*, 4193-4194.

[284] Spackova, N.; Berger, I.; Sponer, J. *J. Am. Chem. Soc.*, **1999**, *121*, 5519-5534.

[285] Kettani, A.; Gorin, A.; Majumdar, A.; Hermann, T.; Skripkin, E.; Zhao, H.; Jones, R.; Patel, D. J. *J. Mol. Biol.*, **2000**, *297*, 627-644.

[286] Liu, H.; Kugimiya, A.; Matsugami, A.; Katahira, M.; Uesugi, S. *Nucleic Acids Res. Suppl.*, **2002**, 177-178.

PUBLICATIONS

1. Anna Lisa Piccinelli, Osmany Cuesta-Rubio, Mariano Barrios Chica, Naheed Mahmood, **Bruno Pagano**, Michele Pavone, Vincenzo Barone and Luca Rastrelli. **Structural revision of clusianone and 7-epi-clusianone and anti-HIV activity of polyisoprenylated benzophenones.** *Tetrahedron* **2005**, 61 (34), 8206-8211.

Abstract

For the first time, the tautomeric pairs of clusianone and 7-epi-clusianone were isolated from the same source, Clusia torresii fruits. An extensive NMR spectroscopic study is described to establish ^{1}H and ^{13}C chemical shift assignments and the C-7 relative configuration of these epimers and to clarify contradictory NMR spectroscopic data previously reported. Quantum mechanical computations than pointed out the relationship between indirect coupling constants and the equilibrium between the B-ring chair and twist-boat forms of the bicyclo-[3.3.1.]-nonane system. Clusianone, 7-epi-clusianone and polyisoprenylated benzophenones 18,19-dihydroxyclusianone, propolone A and nemorosone were screened for their activity against HIV infection in C8166 cells. All compounds inhibited infection with selectivity index values ranging from 2.25 to 15.6. Only clusianone derivatives inhibited infection by binding to viral protein gp120 and prevented its interaction with cellular receptor CD4 as detected by ELISA using recombinant proteins.

2. Luigi Petraccone, **Bruno Pagano**, Veronica Esposito, Antonio Randazzo, Gennaro Piccialli, Guido Barone, Carlo A. Mattia and Concetta Giancola. **Thermodynamics and Kinetics of PNA-DNA Quadruplex-Forming Chimeras.** *J. Am. Chem. Soc.* **2005**, 127 (46), 16215-16223.

Abstract

PNA-DNA chimeras present the interesting properties of PNA, such as the high binding affinity to complementary single-strand (DNA or RNA), and the resistance to nuclease and protease degradation. At the same time, the limitations of an oligomer containing all PNA residues, such as low water solubility, self-aggregation, and low cellular uptake, are effectively overcome. Further, PNA-DNA chimeras possess interesting biological properties as antisense agents. We have explored the ability of PNA-DNA chimeric strands to assemble in quadruplex structures. The rate constant for association of the quadruplexes and their thermodynamic properties have been determined by CD spectroscopy and differential scanning calorimetry (DSC). Thermal denaturation experiments indicated higher thermal and thermodynamic stabilities for chimeric quadruplexes in comparison with the corresponding unmodified DNA quadruplex. Singular value decomposition analysis (SVD) suggests the presence of kinetically stable intermediate species in the quadruplex formation process. The experimental results have been discussed on the basis of molecular dynamic simulations. The ability of PNA-DNA chimeras to form stable quadruplex structures expands their potential utility as therapeutic agents.

3. Pompea Del Vecchio, Paola Carullo, Guido Barone, **Bruno Pagano**, Giuseppe Graziano, Alessio Iannetti, Renato Acquaviva, Silvestro Formisano. **Conformational stability and DNA binding energetics of the rat thyroid transcription factor 1 homeodomain.** *Proteins*, **2007**, 70 (3), 748-760.

Abstract

The conformational stability of the rat thyroid transcription factor 1 homeodomain, TTF-1HD, has been investigated by means of circular dichroism (CD) and differential scanning calorimetry (DSC) measurements at pH 5.0 as a function of KCl concentration. Thermal unfolding of TTF-1HD is a reversible two-state transition. The protein is not stable against temperature, showing a denaturation temperature of 32 degrees C in the absence of salt and 50 degrees C at 75 mM KCl. The binding energetics of TTF-1HD to its target DNA sequence has been characterized by means of isothermal titration calorimetry (ITC) measurements, complemented with CD data. At 25 degrees C, pH 5.0 and 75 mM KCl, the binding constant amounts to 1.5 x 10^{8} M^{-1} and the binding enthalpy change amounts to -41 kJ mol^{-1}. The process is enthalpy driven, but also the entropy change is favorable to complex formation. To gain a molecular level understanding of the

interactions determining the association of TTF-1HD to the target DNA sequence structural information would be requested, but it is not yet available. Therefore, structural models of two complexes, TTF-1HD with the target DNA sequence and TTF-1HD with a modified DNA sequence, have been constructed by using as a template the NMR structure of the complex between NK-2 HD and its target DNA, and by performing molecular dynamics simulations 3.5 ns long. Analysis of these models allows one to shed light on the origin of the DNA binding specificity characteristic of TTF-1HD.

4. **Bruno Pagano** and Concetta Giancola. **Energetics of quadruplex-drug recognition in anticancer therapy.** *Curr. Cancer Drug Targets*, **2007**, 7 (6), 520-540.

Abstract
Immortality of tumour cells is strictly correlated to telomerase activity. Telomerase is overexpressed in about 85% of tumour cells and maintains telomere length contributing to cell immortalisation, whereas in somatic cells telomeres progressively shorten until cell death occurs by apoptosis. Different drugs can promote telomeric G-rich overhangs which fold into quadruplex structures that inhibit telomerase activity. Detailed studies on drug-quadruplex complexes are essential to understand quadruplex recognition and address drug design. This review will discuss the energetic aspects of quadruplex-drug interactions with a particular attention to physico-chemical methodologies.

5. Angela Bisio, **Bruno Pagano**, Alessia Romussi, Olga Bruno, Nunziatina De Tommasi, Giovanni Romussi and Carlo A. Mattia. **Relative Stereochemistry of a Diterpene from Salvia cinnabarina.** *Molecules*, **2007**, 12 (10), 2279-2287.

Abstract
The relative stereochemistry of 3,4-secoisopimara-4(18),7,15-triene-3-oic acid, a diterpenoid with antispasmodic, hypotensive and antibacterial activities isolated from Salvia cinnabarina, was determined by an X-ray diffraction analysis of a single crystal of a suitable crystalline derivative.

6. Ada Virno, Luciano Mayol, Andres Ramos, Franca Fraternali, **Bruno Pagano**, Antonio Randazzo. **Structural insight into the hTERT intron 6 sequence d(GGGGTGAAAGGGG) from 1H-NMR study.** *Nucleosides Nucleotides Nucleic Acids*, **2007**, 26 (8-9), 1133-1137.

Abstract
The interest in DNA quadruplex structures has been fueled by the recognition that telomeres, the 3' single stranded guanine-rich overhangs found at the termini of chromosomes, are likely to form G-tetrads type structures important in cell senescence and cancer. In addition to their presence in telomeres, where they may play a role in maintaining the stability and integrity of chromosomes, guanine-rich regions are found in other region of the genome, amongst these is intron 6 of hTERT a gene codifying for the enzyme telomerase. Interestingly, the formation of G-quadruplexes in this region is involved in the down-regulation of telomerase activity caused by an alteration of the hTERT splicing pattern. Therefore, we have analyzed several sequences of that intron by ^{1}H-NMR and CD spectroscopy, and we have found that the sequence d(GGGGTGAAAGGGG) is able to fold in a single well-defined antiparallel quadruplex structure consisting of four G-tetrads, possessing a twofold symmetry, and containing four Gs in a syn glycosidic conformation.

7. **Bruno Pagano**, Carlo A. Mattia, Ada Virno, Antonio Randazzo, Luciano Mayol, Concetta Giancola. **Thermodynamic analysis of Quadruplex DNA-drugs interaction.** *Nucleosides Nucleotides Nucleic Acids*, **2007**, 26 (6), 761-765.

Abstract

This work studies the binding properties of distamycin and its carbamoyl analog, containing four pyrrole units, with the [d(TGGGGT)]$_4$ quadruplex by means of isothermal titration calorimetry (ITC). Analysis of the ITC data reveals that drug/quadruplex binding stoichiometry is 1:1 for both interactions and that distamycin analog gives approximately a 10-fold increase in the quadruplex affinity.

8. Luigi Martino, Ada Virno, **Bruno Pagano**, Antonella Virgilio, Simone Di Micco, Aldo Galeone, Concetta Giancola, Giuseppe Bifulco, Luciano Mayol and Antonio Randazzo. **Structural and Thermodynamic Studies of the Interaction of Distamycin A with the Parallel Quadruplex Structure [d(TGGGGT)]4.** *J. Am. Chem. Soc.*, **2007**, 129, 16048-16056.

Abstract

The complex between distamycin A and the parallel DNA quadruplex [d(TGGGGT)]$_4$ has been studied by ^{1}H NMR spectroscopy and isothermal titration calorimetry (ITC). To unambiguously assert that distamycin A interacts with the grooves of the quadruplex [d(TGGGGT)]$_4$, we have analyzed the NMR titration profile of a modified quadruplex, namely [d(TGG$_{Me}$GGT)]$_4$, and we have applied the recently developed differential frequency-saturation transfer difference (DF-STD) method, for assessing the ligand-DNA binding mode. The three-dimensional structure of the 4:1 distamycin A/[d(TGGGGT)]$_4$ complex has been determined by an in-depth NMR study followed by dynamics and mechanics calculations. All results unequivocally indicate that distamycin molecules interact with [d(TGGGGT)]$_4$ in a 4:1 binding mode, with two antiparallel distamycin dimers that bind simultaneously two opposite grooves of the quadruplex. The affinity between distamycin A and [d(TGGGGT)]$_4$ enhances (approximately 10-fold) when the ratio of distamycin A to the quadruplex is increased. In this paper we report the first three-dimensional structure of a groove-binder molecule complexed to a DNA quadruplex structure.

9. **Bruno Pagano**, Luigi Martino, Antonio Randazzo and Concetta Giancola. **Stability and Binding Properties of a Modified Thrombin Binding Aptamer.** *Biophys. J.*, **2008**, 94 (2), 562-9.

Abstract

Aptamer-based drugs represent an attractive approach in pharmacological therapy. The most studied aptamer, thrombin binding aptamer (TBA), folds into a well-defined quadruplex structure and binds to its target with good specificity and affinity. Modified aptamers with improved biophysical properties could constitute a new class of therapeutic aptamers. In this study we show that the modified thrombin binding aptamer (mTBA), (3')GGT(5')-(5')TGGTGTGGTTGG(3'), which also folds into a quadruplex structure, is more stable than its unmodified counterpart and shows a higher thrombin affinity. The stability of the modified aptamer was investigated using differential scanning calorimetry, and the energetics of mTBA and TBA binding to thrombin was characterized by means of isothermal titration calorimetry (ITC). ITC data revealed that TBA/thrombin and mTBA/thrombin binding stoichiometry is 1:2 for both interactions. Structural models of the two complexes of thrombin with TBA and with mTBA were also obtained and subjected to molecular dynamics simulations in explicit water. Analysis of the models led to an improvement of the understanding of the aptamer-thrombin recognition at a molecular level.

10. **Bruno Pagano**, Ada Virno, Carlo A. Mattia, Luciano Mayol, Antonio Randazzo and Concetta Giancola. **Targeting DNA quadruplexes with distamycin A and its derivatives: an ITC and NMR study.** *Biochimie*, **2008**, 90, 1224-32.

Abstract

The use of small molecules that bind and stabilize G-quadruplex structures is emerging as a promising way to inhibit telomerase activity in tumour cells. In this paper, isothermal titration calorimetry (ITC) and ^{1}H-NMR studies have been conducted to examine the binding of distamycin A and its two carbamoyl

derivatives (compounds 1 and 2) to the target $[d(TGGGGT)]_4$ and $d[AG_3(T_2AG_3)_3]$ quadruplexes from the *Oxytricha* and human telomeres, respectively. The interactions were examined using two different buffered solutions containing either K^+ or Na^+ at a fixed ionic strength, to evaluate any influence of the ions present in solution on the binding behaviour. Experiments reveal that distamycin A and compound 1 bind the investigated quadruplexes in both solution conditions, conversely compound 2 appears to have a poor affinity in any case. Moreover, these studies indicate that the presence of different cations in solution affects the stoichiometry and thermodynamics of the interactions.

11. **Bruno Pagano**, Carlo A. Mattia, Luigi Cavallo, Seiichi Uesugi, Concetta Giancola and Franca Fraternali. **Stability and cations coordination of DNA and RNA 14-mer G-quadruplexes: a multiscale computational approach.** *J. Phys. Chem. B*, accepted.

Abstract
Molecular dynamics simulations have been used to study the differences between two DNA and RNA 14-mer quadruplexes of analogous sequences. Their structures present a completely different fold: DNA forms a bimolecular quadruplex containing antiparallel strands and diagonal loops; RNA forms an intrastrand parallel quadruplex containing a G-tetrad and an hexad, which dimerizes by hexad stacking. We used a multiscale computational approach combining classical Molecular dynamics simulations and density functional theory calculations to elucidate the difference in stability of the 2-folds and their ability in coordinating cations. The presence of 2'-OH groups in the RNA promotes the formation of a large number of intramolecular hydrogen bonds that account for the difference in fold and stability of the two 14-mers. We observe that the adenines in the RNA quadruplex play a key role in conserving the geometry of the hexad. We predict the cation coordination mode of the two quadruplexes, not yet observed experimentally, and we offer a rationale for the corresponding binding energies involved.