UNIVERSITAT DE BARCELONA

Studies on the formation of i-motif structures at neutral pH. Use of cytidine analogues and importance of minor groove tetrads on mini i-motifs stabilization

Bartomeu Mir Morro

PROGRAMA DE DOCTORAT EN QUÍMICA ORGÀNICA

STUDIES ON THE FORMATION OF i-MOTIF STRUCTURES AT NEUTRAL pH. USE OF CYTIDINE ANALOGUES AND IMPORTANCE OF MINOR GROOVE TETRADS ON MINI i-MOTIFS STABILIZATION.

Manuscript submitted by:

BARTOMEU MIR MORRO

Directed and supervised by:

NÚRIA ESCAJA SÁNCHEZ

Tutor:

ENRIQUE PEDROSO MÜLLER

Departament de Química Inorgànica i Química Orgànica.

Secció de Química Orgànica.

Facultat de Química. Universitat de Barcelona.

AGRAÏMENTS

Núria, em sento molt afortunat d'haver-te tingut com a directora de tesi. Durant aquests anys he aprés i crescut molt. T'agraeixo infinitament la confiança, la proximitat i la motivació que m'has demostrat a diari.

Gracias también a ti, Carlos, por la predisposición a la ayuda, el consejo y el buen humor en todo momento. Gracias también a Irene, Miguel y Isra por la complicidad y la inmejorable compañía en estancias y congresos

Alguns dels resultats fonamentals d'aquesta tesi no haguessin estat possibles sense la col·laboració amb el grup del Professor M. Orozco. Per això us vull agrair Albert, Montse i Diana, els vostres esforços abocats en aquesta tesi

Gràcies també a tots els cops de mà que he rebut sempre, i sense pregar, al personal de la facultat: Josep de la sala d'aparells, Irene i Laura del MALDI i a tota la gent del magatzem i secretaria.

Als membres del grup, Anna, Enrique, Jordi i Vicente, gràcies per fer-me sentir sempre benvingut i per les vostres continues mostres de suport durant tots aquests anys.

A tots els companys de laboratori i de la facultat, dels quals he obtingut sense ben bé esperar-ho algunes amistats molt valuoses. Dels primers als últims, espero no deixar-ne cap. Albert, Àlex, Anna, Berta, Clément, Jordi, Josep, Natàlia, Omar i Pau. Gràcies per tots els bons moments i per mostrar-me el camí en la recerca. Us admiro.

Tinc molta sort de comptar amb molts bons amics. Al ghetto, un bocí fonamental de Mallorca a Barcelona, companys de pis envejables: Anna, Elena, Maria, Marta i Rafel. També al meu nucli dur, els de sempre, custodiant la tranquil·litat illenca: Marta, Sara i Toni. Vos estim a tots igual.

Per mon pare i ma mare no bastarien aquestes 300 pàgines. Vos ho dec tot. Gràcies per un amor incondicional.

A tu, Maria, rebel i consentida, gràcies per ser tan divertida.

TABLE OF CONTENTS

ABREVIATIONS

A	Adenine
Abs	Absorbance
APS	Ammonium persulfate
bp	Base pair
C	Cytosine
CD	Circular dichroism
CNE	2-Cyanoethyl
CPG	Controlled pore glass
DNA	Deoxyribonucleic acid
G	Guanine
HPLC	High performance liquid chromatography
HT	Hoogsteen
MALDI	Matrix assisted laser desorption ionization
^{m}C	5-Methylcytosine
mdeg	Millidegrees
MGT	Minor groove tetrad
MSNT	1-(mesitylene-2-sulfonyl)-3-nitro-1,2,4-triazole
NMR	Nuclear magnetic resonance
NOE	Novel Overhauser effect
NOESY	Novel Overhauser effect spectroscopy
Nt	Nucleotide
OD	Optical Density
p	Phosphate
PAGE	Polyacrylamide gel electrophoresis

PG Protecting group

pH_T Transitional pH

psC Pseudoisocytidine

RNA Ribonucleic acid

SPOS Solid phase oligonucleotide synthesis

T Thyimine

TCA Trichloroacetic acid

tC^o 1,3-Diaza-2-oxophenoxazine

TEA Triethylamine

TEAA Triethylammonium acetate

TEMED Tetramethylethylenediamine

T_m Melting temperature

TOCSY Total correlation spectroscopy

TOF Time of flight

t_R Retention time

U Uracil

UV Ultraviolet

WC Watson-Crick

1. INTRODUCTION

1.1. DNA NON-CANONICAL STRUCTURES

Undoubtedly, the structural elucidation of DNA by Watson and Crick in 1953,[1] provided by different X-ray diffraction experiments performed by Franklin[2] and Wilkins,[3] is one of the top highlights in modern science history. The finding that DNA folds into a double stranded right-handed helix, triggered the discovery and explanation of several vital biological processes that allowed the scientific community to depicture the secrets of life itself.

It is widely accepted that the canonical double helix is the form that is normally adopted by DNA molecules *in vivo*. Interestingly, from the unwinding of the DNA strands and under supercoiling conditions that can take place during certain cellular processes, several structures different from the double helix have been observed.[4] These non-canonical structures are specially adopted by repetitive sequences and have been found to be strategically distributed throughout the genome. Their formation has been proven to induce genetic instability and to cause DNA breakage, which might ultimately cause human diseases.[5] The description of this variety of structures has confirmed the intrinsic flexible nature of DNA in terms of base pairing, molecularity, stability and recognition.

DNA is best known for being responsible of the storage, transmission and expression of genetic information in all living cells. Eukaryotic organisms store large amounts of compacted DNA in the nucleus of their cells, organized in macrostructures called chromosomes. During cell division, DNA replication gives rise, via complementary base pairing, to a complete identical set of chromosomes for each daughter cell. Coding genetic material may undergo metabolic cellular processes such as transcription and translation that lead, to the synthesis of specific RNA molecules and proteins. These cellular processes are of high complexity and require the implication a large variety of proteins. However, it is important to mention that nucleic acids roles are not only related to genes and protein synthesis but in fact, mostly to non-coding purposes such as regulation or structural functions. Non-canonical structures have been suggested to be involved in some of these functions.[6]

Thus, it is unquestionable how the knowledge on DNA structure is crucial for properly understanding its biological function. For this reason, one of the hot topics for researchers has been, during the last decades, to put light on the specific biological roles that non-canonical DNA structures might play in the cell. Moving one step forward, due to their biological implications, non-canonical structures have been recently considered as potential therapeutic targets, to ultimately regulate the cellular processes in which they might be involved. To this aim, a huge amount of work has been performed in the field of structural biology. In the following sections of this introduction, a classification of the different DNA structures that have been reported, with special attention to their possible biological and therapeutic implications, will be done. An extended and more detailed review will be done for quadruplex structures, specially focusing on i-motif structures.

1.1.1. DUPLEXES

1.1.1.1. RIGHT-HANDED CANONICAL FORMS: A-DNA & B-DNA

The structural flexibility of DNA was already reported in the first of X-ray diffraction. [2,3] Under different experimental conditions, the crystallographers observed two different right-handed double helices that are represented in *Figure* 1-1.

The B-DNA form is the one described by Watson and Crick and the one that is preferably adopted *in vivo* by DNA.[1] Its formation was first observed from X-ray experiments under high humidity conditions. The structure can be described as a right-handed double helix, where the association of the strands takes place by hydrogen bond interactions between their nucleobases in an antiparallel manner. Watson and Crick described the complementarity between purine and pyrimidine nucleobases and thus, canonical A:T and G:C base pairs are named after them as Watson-Crick (WC) base pairs. The base pairs are placed in the center of the helix and their plane is almost perfectly perpendicular to the axis of the helix. The sugar-phosphate backbone is left facing the exterior of the helix. The B-DNA helix presents 10.5 nucleotides per turn, with a C2'-*endo* sugar puckering and an *anti* configuration of the glycosidic bonds. Overall, these characteristics prevent the strands from being symmetrically opposite to each other and lead to the formation of two well differentiated grooves: a major groove (22 Å wide) and a minor groove (12 Å wide). These grooves are of dramatic importance for the recognition of the structure by proteins and small molecules.

The other right-handed helix observed by X-ray diffraction in 1953 was obtained under dehydrating and high ionic strength conditions, which are common conditions used for crystallization. Very interestingly, this form was later found to be adopted by double stranded RNA fragments and by DNA-RNA hybrids.[7] Named A-DNA, or A form, this helix also presents two well-differentiated grooves, antiparallel strands, WC base pairing and an *anti* glycosidic bond configuration, but differs from the B form in the sugar conformation and in the overall shape of the helix. A-DNA helices present 11 nucleotides per turn, with C3'-*endo* conformation of the sugar ring. The base pairs are inclined approximately 20º toward the axis of the helix, conferring a more compacted and wide shape to the structure compared to that of the B form. These differences also affect the shape of the grooves, resulting in a narrow and deep major groove and in a wide and flattened minor groove.

These different helical features are of crucial relevance in terms of recognition by proteins and small molecules between dsDNA and dsRNA (or DNA-RNA hybrids). Transitions between A and B forms have been reported to occur by modulating the medium conditions, and to take place *in vivo* in DNA-protein interactions.[8]

1.1.1.2. LEFT-HANDED HELIX: Z-DNA

Serendipity wanted that, back in 1979, the crystal structure from the DNA fragment d(CGCGCG), the first to ever be solved, afforded an unexpected left-handed double helix.[9] Named Z-DNA after the zig-zag-like shape described by its sugar-phosphate backbone, this structure presents, contrary to the A and B forms, a dinucleotide as a repetitive unit. Sequentially restricted to purine-pyrimidine repeats, Z-DNA is a left-handed antiparallel double helix, with WC base pairing, that exhibits significant conformational differences compared to the B-DNA. Whilst pyrimidine nucleotides maintain its C2'-*endo* sugar puckering and *anti* conformation of the glycosidic bond, purines, which are generally more prone to it, adopt C3'-*endo* and *syn* configurations. This alternation in nucleotide configuration is the reason behind the characteristic zig-zag-like shape of Z-DNA. The resultant helix, an example of which is shown in *Figure* 1-1, is slightly thinner than the B form, with 12 nucleotides per turn and a unique narrow and deep groove.

Although the formation of Z-DNA can be induced by increasing the ionic strength of the medium in order to prevent the repulsion between the phosphate groups,[10] it was found that negative supercoiling conditions favor its formation in physiological medium.[11] Thus, one can imagine a dynamic transition from the B to Z form, *in vivo*, as a conformational rearrangement in supercoiled DNA. Subsequently, the study of its biological function is conditioned to its transitional formation. Interestingly, sequences that match the requirements for the formation of Z-DNA have been found close to promoter regions of the genome, indicating a possible role of this non-canonical structure as a transcriptional regulator.[12] Also, highly specific proteins such as ADAR1, have been reported to bind with strong affinity to Z-DNA and Z-RNA,[13] emphasizing its existence and relevance *in vivo*.

1.1.2. PARALLEL DUPLEXES

Parallel orientations were first described in the early 1960s for polyadenylic acid molecules at acidic pH. The formation of this motif was explained from the study of different poly(A) sequences to occur thanks to the association of protonated adenine nucleobases thorough their Hoogsteen side.[14] However, several research groups have reported a wide range of base pairing patterns, depending on sequence, pH and stabilizing agents that allow the formation of parallel duplexes. Duplexes with Hoogsteen A:T and G:C^+ base pairs show better stabilities under acidic conditions due to protonation requirement of cytidine.[15] Poly[d(A:T)] sequences adopt a parallel right-handed double helix stabilized by reverse Watson-Crick base pairs. Its helical features are similar to those of the antiparallel B form, with perpendicular orientations of the base pairs towards the axis of the helix, south conformations of the sugar rings and two equivalent grooves (see *Figure* 1-1). Parallel duplexes containing G:G, A:A, C:C^+ and T:T homopairs have also been proposed. Their formation is pH-dependent and occur favorably in d(CGA) repeats.[16] Moreover, parallel orientations have been described in other multi-stranded DNA structures as it will be described in following sections of this introductory chapter.

Specific biological functions of parallel duplexes have not been described, although potentially parallel duplex-forming sequences have been found in strategic regions of the genome and their evolutionary role remains under discussion. Parallel-stranded duplexes might be utilized as a successful template for the formation of triple-stranded DNA structures, which might be of relevant help for biotechnological and antigene and antinsense therapeutic uses.[17]

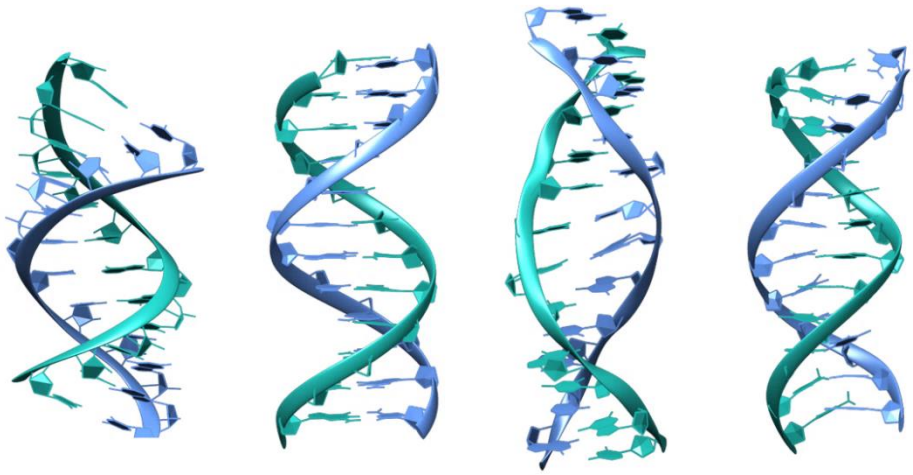


Figure 1-1.Representation of different nucleic acid duplexes, from left to right: A-RNA (PDB code 4KYY), B-DNA (PDB code 2M2C), Z-DNA (PDB code 3IRQ) and DNA parallel duplex (PDB code 1JUU).

1.1.3. STRUCTURES WITH UNPAIRED REGIONS

1.1.3.1. BULGES AND INTERNAL LOOPS

Double helical nucleic acids might exhibit punctual structural disruptions provoked by unpaired nucleobases. These mismatches can be classified depending on if they affect to just one or both strands. While bulges generate a kink in just one of the strands, provoked by the loss of WC base pairing of a certain number of nucleotides, internal loops, which can be symmetrical or asymmetrical, present unpaired nucleotides in both strands of the double helix.

Bulges and internal loops seem to work as a recognition site in the overall tertiary structure of nucleic acids *in vivo*. Of special interest is the case of high order RNA architectures such as HCV IRES (see *Figure* 1-2) or flaviviruses (dengue, yellow fever, etc.), that build their structure as combination of double helical fragments together with bulges, internal loops and stem loops.[18]

1.1.3.2. HAIRPINS AND CRUCIFORMS

Stem loops, also known as hairpins or hairpin loops, are formed from specific sequences known as inverted or mirror repeats. They are formed from a single-stranded nucleic acid containing reverse auto-complementarity and thus, with the ability to fold into a B-like intramolecular helix via WC base pairing joined by an unstructured kink or loop. Their stability is dependent of the length and nature of both the double helix that is newly formed and the stem loop.[19] Cruciforms, on the other hand, are formed by the folding of a fragment of each of the strands of a double helix into a hairpin, resulting in an X-shaped overall structure. Cruciforms are specially favored under superhelical stress conditions.

As bugles and internal loops, hairpins seem to have an important role as recognition sites. Their formation in DNA sequences can arise from two different mechanisms. The first one implies the prior unwinding of the double helix which can occur during gene expression or cell replication. The other one occurs from the excision of two single hairpins from a previously formed cruciform. Hairpins can be also found as building blocks of other nucleic acids structures such as intramolecular triplexes. Hairpins are ubiquitous in RNA structures. Transfer RNA is possibly the most famous example of the importance of these structures due to forming the anticodon that recognizes a codon during translation. Other functions involve RNA structural stabilization and messenger RNA regulation.[20]

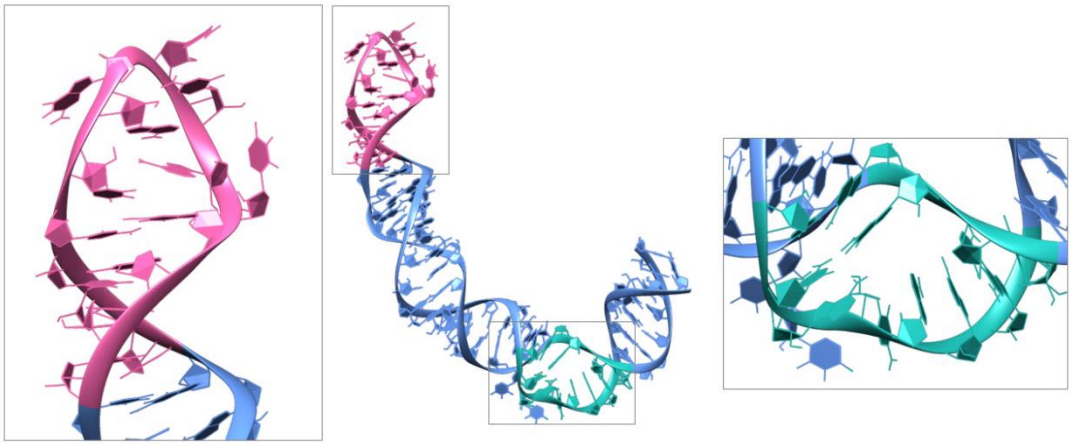


Figure 1-2. Hepatitis C virus IRES subdomain II (center, PDB code 1P5P). Ampliations of the hairpin loop from the subdomain IIb (left) and of the bulge from the subdomain IIa (right).

1.1.4. TRIPLEXES

First described in a 2:1 poly(dU):poly(dA) mixture in 1957,[21] triple-stranded associations of nucleic acids (triplexes) are stabilized by the interaction of WC base pairs and a third nucleotide via their Hoogsteen side. A general scheme for the formation of triplex structures implies the approximation of a single-

stranded oligonucleotide to an already formed WC duplex through its major groove. The formation of triplexes is favored by d(purine:pyrimidine) duplexes since a purine-rich strand is required for the favorable formation of Hoogsteen base pairs. Depending on the sequential nature and the orientation adopted by the third strand, triplex structures are classified into two groups. Pyrimidine or parallel triplexes are formed by the parallel association of a pyrimidine-rich strand with the purine-rich strand of the duplex via Hoogsteen base pair formation. The typical triplets that are formed in parallel triplexes are T-A:T and C^+-G:C (see *Figure* 1-3), with cytidine protonation of the N3 nitrogen atom.[22] On the other hand, purine or antiparallel triplexes occur between a purine-rich strand that is reverse Hoogsteen bonded to a duplex in an antiparallel fashion.[23] The common triplets found in antiparallel triplexes are G-G:C and A-A:T. Inter- and intramolecular triplex formation are both feasible and, although its formation is kinetically limited, triplexes generally show great thermodynamic stability conditioned by nucleotide sequence, length of the incorporated strand, divalent cations such as Mg^{2+}, and pH.[24]

Intramolecular triplexes from biological DNA are reported to be formed from the partial unwinding of the double helix, that is instigated by superhelical stress, and the following fold back of a single strand to form the triplex. Mirror repeats characteristically favor the formation of these motifs. Different nomenclature is applied to differentiate the strand configuration of this motif. H-DNA corresponds to a parallel triplex from the rearrangement of a pyrimidine-rich strand,[25] whereas *H-DNA refers to the antiparallel conformation. The formation of H-DNA *in vivo* is strongly supported specially by studies performed using S1 nuclease, which is selective to single-stranded DNA. It has been suggested that the role of H-DNA is related with gene transcription regulation since potentially triplex-forming sequences have been found in non-coding regions. Specific binding proteins for H-DNA have been reported.[26]

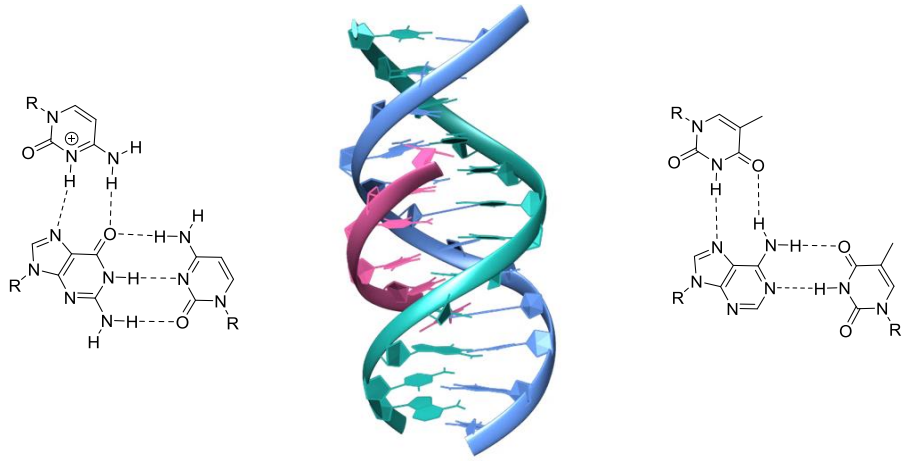


Figure 1-4. Representation of a parallel triplex (PDB code 1BWG) with C^+-G:C (left) and T-A:T (right) triplets.

1.1.5. TETRAPLEXES

1.1.5.1. G-QUADRUPLEX

The aggregation of guanine-rich strands in tetrameric macrostructures was first described by fiber diffraction X-ray experiments in 1962.[27] The authors already proposed that the stabilizing unit of this motif was a cyclic tetrad formed by four guanine residues held together by hydrogen bond interactions through their Hoogsteen side (*Figure* 1-6, center)). The formation of a G4 tetrad requires the incorporation of a monovalent cation to compensate the electrostatic repulsion generated by the guanine O6 oxygen atoms. As in other DNA secondary structures, stacking interactions are crucial for the formation of G-quadruplexes, that must present, at least, two consecutive G4 tetrads stacked one onto another in order to be stable. The stacking disposition of the tetrads leads to an overall four-stranded helical-shaped structure with four well-defined grooves.[28]

Extensive research performed during more than five decades has led to confirm that G-quadruplexes are a very varied structures in terms of molecularity and topology and that a wide range of G-tracts can associate to stably form this motif. The formation of G-quadruplex can be achieved from the association of one, two or four different nucleic acid strands, either DNA or RNA. Remarkably, the stability of RNA G-quadruplexes is enhanced by intra-strand interactions of the 2'-hydroxyl group.

For quadruplex motifs, the construction of the structure may adopt different topologies, depending on the strands directions and the possibility of combining G4 tetrads with canonical WC base pairs. Moreover, the length and the sequence of the loops connecting the guanine runs in intramolecular and dimeric G-quadruplexes, confers even more versatility to the topology of the structure.[28] So-called parallel G-quadruplexes present their four strands in the same orientation and thus must present, for intramolecular and dimeric motifs, a propeller loop connecting the 3'-end of one G-tract to the 5'-end of an adjacent one.[29] Propeller loops are of special significance since their placement, depending on length and sequence, might notably disrupt the topology of the grooves, directly affecting to possible recognition interactions either by proteins or small molecules. In the other hand, antiparallel G-quadruplexes present, at least, one of the strands in an opposite orientation to the others. For antiparallel motifs, two more types of loops can be described. Lateral loops connect adjacent antiparallel G-tracts and their relative position in dimeric motifs may afford either head-to-head or head-to-tail orientations.[30] Diagonal loops connect opposite antiparallel G-tracts and its formation implies an alternate orientation distribution between adjacent runs.[31] Contrary to propeller loops, structures with only lateral or diagonal loops exhibit topologically simpler grooves. Differences between parallel and antiparallel motifs also involve glycosidic bond configurations. For a graphic representation of the previously described concepts see *Figure* 1-5. Whereas for parallel G-quadruplexes all guanine residues adopt an *anti* configuration, a combination of *syn* and *anti* conformations is adopted for each antiparallel motif. Circular dichroism is sensitive to these different configurations and it is used as a very useful tool for G-quadruplex characterization.[28]

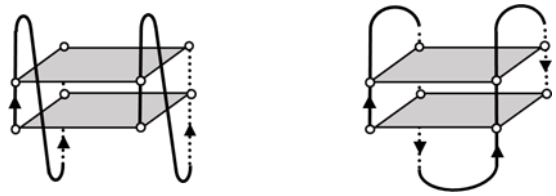


Figure 1-5. Examples of different topologies and loops found in G-quadruplex structures. Dimeric parallel G4 with propeller loops (left) and unimolecular antiparallel G4 with two lateral and one diagonal loop (right).

The stability of G-quadruplexes is strongly conditioned by its topology and other factors. Thermodynamic data points to a greater stability for increasing number of G4 tetrads.[32] The size of the loops effect has also been studied by thermodynamic and NMR experiments concluding that, in general terms, short loops favor parallel orientations while long loops lead to the formation of antiparallel motifs. The topology for intermediate loops is more unpredictable and a large variety of conformations have been reported. Diagonal loops are less likely to be formed by short loops. All nucleobases are compatible with loop formation and their nature affects the stability of the structure depending on potential stacking interactions.[33,34]

As stated above, monovalent cations are indispensable for G-quadruplex formation. K^+ is best known for giving rise to lower energy structures and is placed in the central channel of the structure, equidistantly between two consecutive G4 tetrads. Na+, on the other hand, also affords stable G-quadruplexes by coordinating O6 guanine oxygen atoms either of one (coplanar) or two (as K^+ does) G4 tetrads.[35] Topological alternations by cation change depends on each structure. While some structures suffer a notable modification when in presence of Na^+ or K^+, others remain unaltered.

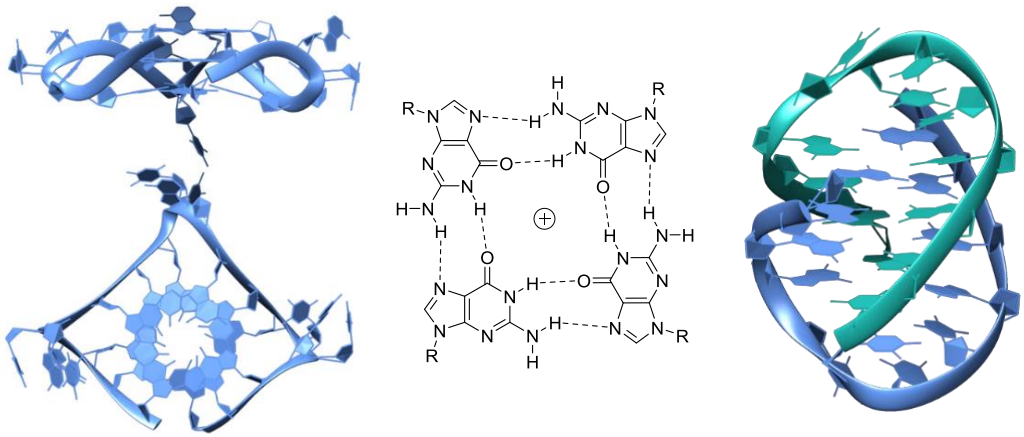


Figure 1-6. Lateral and top views of an intramolecular G-quadruplex (left, PDB code 3UYH). G4 tetrad (center). Dimeric G-quadruplex with diagonal loops (right, PDB code 4R45).

Interestingly, a few years before the first G-quadruplex high-resolution structures were obtained in 1992,[36] guanine-rich telomeric sequences from the immunoglobulin switch region were reported to fold into G-quadruplex structures.[37,38] Consequently, these structures moved from being considered a mere structural casualty to a potentially relevant biological feature. First, in order to consider its formation *in vivo,* G-quadruplex versus duplex formation is an important question to consider. More specifically, one may wonder in which situations DNA is likely to abandon its canonical B form to become single-stranded and thus fulfill the requirement for folding into an intramolecular G-quadruplex. As already mentioned in this introduction, DNA may undergo processes such as replication, transcription or repair in where, as a transitory state, the double helix is unwound. Moreover, certain conditions like superhelical stress[39] or molecular crowding,[40] as well as specific binding proteins,[41] positively contribute to G-quadruplex formation. In addition, since monovalent cations are crucial for its formation, physiological conditions are optimal for G-quadruplex formation.

Computational analyses performed with the consensus sequence $G_nX_{1-7}G_nX_{1-7}G_nX_{1-7}G_n$, revealed that in the human genome, potentially G-quadruplex forming sequences are found, in an oversimplified approximation, over 300k times. Their location is not arbitrary and, remarkably, are highly conserved between different species.[42]

Of special interest for G-quadruplex formation *in vivo* is the case of telomeres, where the highest abundance of G-quadruplex forming sequences is found.[42] Telomeres are nucleic acid and protein complexes located at the ends of the chromosomes of all eukaryotic organisms. Sequential-wise, telomeres are formed by tandem repeats of 5-10 kb, which in the case of all vertebrates present as repeating unit the fragment d(TTAGGG). Furthermore, the 3'-end of the telomeres presents 100-200 nucleotides single-stranded overhang rich in guanines which is even more prone to the formation of G-quadruplexes. The role of telomeres is as capping units, preventing the identification of the chromosomal ends by the cell as a repairing site, which would lead to the merging of different chromosomes. In a significantly high percentage of human cancers (80-85%), telomerase, the enzyme responsible for the preservation of telomeres after replication processes, is overexpressed. It is suggested that the formation of G-quadruplex structures blocks the activity of nucleases and telomerases, giving proof of a specific and crucial role for a non-canonic DNA structure.[43]

Telomeric RNA, which is transcribed from the complementary cytosine-rich sequence of the potentially forming G-quadruplex sequence, also has the potential to fold into a G-quadruplex. Telomeric repeat-containing RNA (TERRA) origins from the subtelomeric region and presents as a repeating unit the fragment r(UUAGGG). TERRA levels are related to the regulation and the stability of telomeres, indicating another possible direct function of G-quadruplexes.[44,45]

Protein interactions also are of notable interest for depicturing the specific mechanism of G-quadruplex formation and function *in vivo*. Several *in vitro* studies support that chaperones and specially helicases such as WRN and BLM interact with G-quadruplexes in telomeric regions. Human telomeric binding protein TRF2 has also been found to bind both DNA and TERRA G-quadruplexes.[41]

Notably, the *in vivo* visualization of G-quadruplexes was achieved in human cells using the BG4 antibody, which showed the best performance in an *in vitro* screening experiment with affinity constant values from 0.5 to 2 nM.[46] The antibody was proven to recognize and bind different G-quadruplex structures from different sequences and morphology and to specifically bind the quadruplex motif against other nucleic acid secondary structures. The incubation of BG4 with other amplifying fluorescence signal antibodies allowed the visualization of the G-quadruplex-BG4 complex in different cell lines. The same research group demonstrated that BG4 is useful for the visualization of both DNA and RNA G-quadruplexes.[47] Interestingly, when incubated with chromosomes, only 25% of foci corresponded to telomeric regions, indicating that endogenous G-quadruplex structures are formed largely in non-telomeric regions.

Human genes promoter regions have been described to be also rich in G-quadruplex forming sequences.[42] They are found more abundantly in oncogenes (*e.g.*, MYC, KRAS, VEGF or PDGF-A) and regulatory genes. Several gene expression experiments have revealed that enhancing or stabilizing the formation of G-quadruplexes by different means results in an alteration of the gene expression by impairing initiation of transcription.[48] 5'-end untranslated regions (5'-UTR) also showed, computationally, the potential to adopt RNA G-quadruplex structures. Their role has been reported to be the regulation, either by impediment or activation, of the translation of certain oncogenes.[49]

G-quadruplexes being directly involved in many important biological processes, has led many investigators to focus their research in the design of small molecules to specifically bind this structural motif. The therapeutic potential behind targeting G-quadruplexes specially comes from the possibility of actively use their ligands for cancer treatment by interfering with telomeres and oncogenes.[50] An acridine derivative, BRACO-19,[51] and a porphyrin compound, TMPyP4,[52] are possibly two of the best-known specific G-quadruplex ligands that have been reported to show anticancer activity. Interestingly, pyridostatin (PDS) has been reported to inhibit cell growth in human cells by trapping G-quaruplex structures in telomeres.[53] Moreover, PDS derivative carboxyPDS gives proof for a specific G-quadruplex binder with exclusive affinity for RNA G-quadruplexes.[54]

In summary, many prove of the strategic and specific roles played by G-quadruplexes *in vivo* have been enlightened during the last years. With the increasing functional evidences of G-quadruplexes in telomeres and in biological processes such as transcription, translation and replication, , it is easy to understand why this non-canonical DNA structure has received so much attention in last few years.

1.1.5.2. i-MOTIF

Complementary to relevant G-rich sequences, C-rich sequences also have the potential to fold into a very unique tetraplex structure named intercalated (i-) motif. Remarkably, the structural features of G-quadruplexes and i-motifs are very different. Such structural differences between the two tetraplexes

arise from the different repetitive units responsible for the stabilization of both structures. While G4 tetrads stacking stabilizes G-quadruplexes, i-motif structures result from the intercalation of hemiprotonated cytidine-cytidine ($C:C^+$) base pairs. The first evidence for the formation of $C:C^+$ base pairs was reported in 1963 from X-ray diffraction experiments of polycitidylic acid.[55] However, it was not until thirty years later that it was observed that the actual structure of an i-motif consists of a DNA tetraplex stabilized by $C:C^+$ base pairs at acidic pH.[56]

The relevance in biology of i-motif structures has remained under question since its discovery. The special requirements for its formation, led the scientific community to consider the i-motif as a structural oddity, focusing mainly on its nanotechnological applications[57] and ignoring its potential roles *in vivo*. However, during the last two decades, increasing evidences such as i-motif formation in centromeres,[58,59] telomeres[60] or in oncogene promoters,[61,62] have culminated, very recently, with the proposal of their observation *in vivo*.[63,64]

1.1.5.2.1. Structure and stability.

In 1993, Gehring, Leroy and Guéron described, that the hexamer d(TCCCCC) self-associates to form a tetrameric i-motif.[56] Following studies afforded the characterization of numerous i-motif structures concluding that the i-motif can be formed from the association of four, two or a single nucleic acid strand. Generally, the sugar ring and the glycosidic bond of the nucleotides in i-motif structures adopt C3'-*endo* and *anti* conformations, which is theoretically compatible with the formation of RNA i-motif. The i-motif overall structure can be described as the antiparallel association of two parallel duplexes held together by $C:C^+$ base pairs. The association of the two parallel duplexes occurs in a manner that one $C:C^+$ base pair from one of the duplexes is placed intercalated between two $C:C^+$ base pairs of the other duplex. The intercalation of $C:C^+$ base pairs gives rise to a very short distance (~3 Å) between base pairs and renders a right-handed helix with a helical twist angle significantly smaller than that of the B-DNA double helix (~12-20º and 36º, respectively).[65,66] Two wide major grooves and two narrow minor grooves can be well-differentiated in i-motif structures. The sugar-phosphate backbones of the antiparallel strands that make up the narrow grooves are very close to each other (5.9 Å) and produce a destabilizing effect due to negative electrostatic interactions. Interestingly, contacts between sugar rings of different strands are responsible for balancing this destabilizing effect.[67] These sugar-sugar interactions can be classified, because of the antiparallel orientation of the strands, in two types. Face-to-face contacts take place through the 5'-side (O4' oxygens facing each other) of the deoxyriboses whereas back-to-back contacts occur between the 3'-side of the sugar rings.

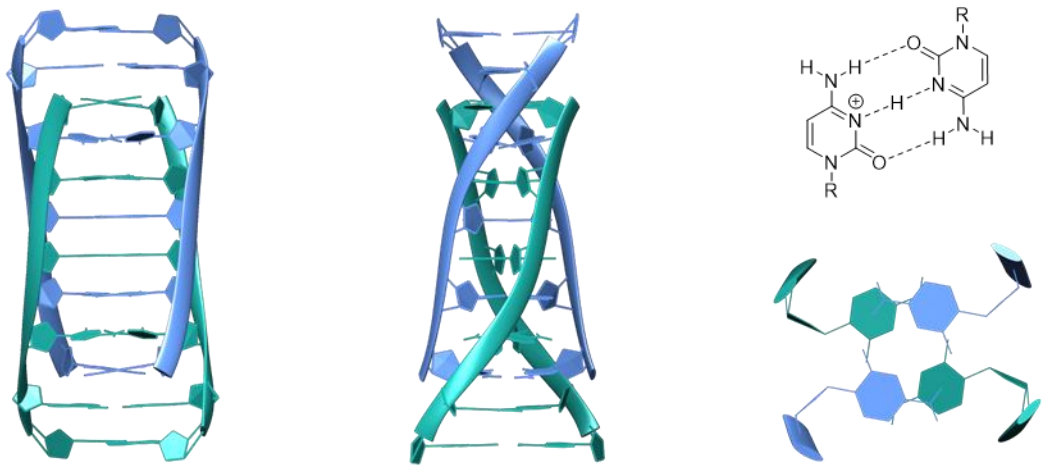


Figure 1-7. Tetrameric i-motif wide (left) and narrow (center) grooves (PDB code 1YBR). Hemiprotonated C:C$^+$ base pair (top right). Overlapped C:C$^+$ base pairs (bottom right).

For non-canonical DNA structures to exist and have relevant *in vivo* activity, they must, first and most importantly, be stable under physiological conditions. Since i-motif formation was initially thought to be restricted to acidic conditions due to protonation requirement of deoxycytidine (pK_a = 4.2 – 4.8),[68] researchers unheeded its biological compatibility. It is worth to mention, however, that during certain processes, the pH of the cell can locally decrease, favoring, theoretically, the formation of i-motif structures. However, wide-ranging research on i-motifs has put light on the different factors that affect the stability of the motif, showing that its formation *in vitro* is possible at neutral and even slightly basic pH.[69] Thus, a busy research line in the i-motif field has tracked the idea of obtaining highly stable structures to certify its feasible formation *in vivo*. In the following sections, a detailed review on the stability of the motif will be performed, from intrinsic sequential characteristics to media conditions, and paying special attention to chemical modifications, which significantly assisted to better understand the clues behind the formation of i-motif structures.

1.1.5.2.1.A. C:C+ stack

The C:C$^+$ base pairs are crucial for i-motifs structures since they are responsible for zipping together the two parallel duplexes. The hydrogen bond pattern of a C:C$^+$ base pair takes place between the amino H42 hydrogen and the carbonyl O2 oxygen atoms. An extra hydrogen bond is formed between the imino N3 nitrogen atoms of the two cytidine residues with a shared H$^+$ atom. The distance between the two N3 nitrogen atoms oscillates between 2.6 and 2.8 Å.[66] The base-pairing energy of a single C:C$^+$ base pair is higher than that of canonical G:C WC base pairs (169.9 kJ/mol and 96.6 kJ/mol, respectively),[70] pointing to an important stability of the hemiprotonated base pair.

The intercalative nature of i-motifs gives rise to interesting properties that are worth to remark. First, and contrary to what happens with most of nucleic acid structures, the stabilizing stacking interactions between adjacent base pairs within the i-motif structure are significantly less important due to poor overlapping among nucleobases.[67] Furthermore, due to the short distance between base pairs, and in order to minimize the repulsive interactions provoked by the stacked positive charges of the hemiprotonated base pairs, computational studies showed that the proton in $C:C^+$ base pairs is asymmetrically shared, placed in the plane of the base pair but at the longest distance from adjacent protons of neighbor $C:C^+$ base pairs.[71] Finally, two different topologies depending on the $C:C^+$ base pairs disposition can be observed in i-motif structures. The 5'-E topology presents at the ends of the stack the $C:C^+$ base pairs corresponding to the 5'-end of the strand/s. On the other hand, in the 3'-E topology the terminal $C:C^+$ base pairs correspond to the 3'-end of the strand/s. In a rough simplification, 3'-E topologies, which present extended sugar-sugar narrow groove interactions, are more stable than 5'-E topologies.[67]

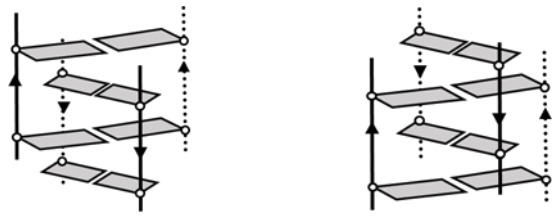


Figure 1-8. 3'E- (left) and 5'E- (right) topologies of i-motif structures.

The number of consecutive $C:C^+$ base pairs that form the structural core of the i-motif has been proven to have a direct influence on its stability. Several studies have delved in this question, concluding that, in a general trend, longer C-tracts afford more stable structures. Moreover, systematic research performed by different groups points towards the same conclusion.[72,73] Utilizing the construction $d(C_nT_3)_3C_n$, both studies observed an increase in both melting temperature (T_m) and transitional pH (pH_T) values from n = 2 (pH_T ~ 6.0) to n = 5 (pH_T ~ 7.2). The tendency for longer constructions is not maintained although no significant decrease in stability is observed. Also, for n values higher than 5-6, two different i-motif species seem to coexist, leading to melting curves with two turning points and high hysteresis.

A similar systematic study performed by Burrow's group with more flexible constructions based on polydeoxicytidine homo-oligonucleotides, revealed an interesting trend by which can be observed that $dC_{(4n-1)}$ sequences afforded the most stable structures for n values between 4 and 7.[74] Shorter nor longer oligonucleotides follow this rule and much longer (> dC_{40}) sequences do not fold into i-motif. In order to control the length of the $C:C^+$ core, researches strategically introduced thymidine residues in

the polydeoxycytidine sequences presenting greater stabilities. Results from this experiment showed lower stabilities than the original oligonucleotides, but remarkably proving that even numbers of $C:C^+$ stacked base pairs afford more stable motifs.

Despite these results, there exist some recent examples that report the formation of highly stable i-motif structures with fewer $C:C^+$ base pairs.[75] The particular case of mini i-motif structures with only two $C:C^+$ base pairs at the core of the structure was proposed by our research group and will be later extended in the discussion section.[76]

1.1.5.2.1.B. Connecting loops

Intramolecular and dimeric i-motifs contain loops at the ends of C-tracts that might notoriously affect the stability of the motif. This effect is not only produced by the length of the loop, but also provoked by the nature of the nucleobases capping the $C:C^+$ stack. Traditionally, depending on the number of nucleotides, connecting loops in i-motif structures can be classified into *class I* or *class II* loops.[77] Whereas *class I* loops comprise loops formed by 2-4 nucleotides, *class II* loops present longer sequences. In an oversimplified approximation, it is generally accepted that *class II* loops confer higher stability to the structure since they present more potentially stabilizing interactions. However, systematic studies based on correlating the thermal and pH stability of the i-motifs formed by $(C_{3,4}T_x)_3C_{3,4}$ constructions, with the length of thymidine loops, showed a clear destabilizing trend for increasing number of thymidine residues.[78] Moreover, recent studies reported highly stable structures with *class I* loops that suggest that this generalization might be misleading and that the actual stabilizing potential within i-motif loops relies in its nucleotide sequence.[79] Interestingly, longer loops with the capability to fold into secondary structures different that the i-motif (*e.g.* hairpin loops), are hypothesized to have a synergistic effect on the stabilization of the i-motif.[80]

Therefore, it seems relevant to consider how the nature of the nucleotides in the loops of i-motif structures influences its stability.[81,82] As mentioned in the previous section, substitution of cytidine residues by thymidine in poly-dC oligonucleotides provokes a destabilization of the motif.[74] Other more systematic study considering the effect of the sequence of the loops on the formation and stability of i-motif structures has been recently performed leading to the same conclusion. When strategically replacing cytidine residues from dC_{15}, dC_{18} and dC_{21} oligonucleotides by T, A or G, in order to obtain structures with $T_{1\text{-}3}$, $A_{1\text{-}3}$ and $G_{1\text{-}3}$ connecting loops, all substitutions led to less stable i-motifs. Notably, greater destabilizations were observed when introducing purine residues in the structure.[73]

Association of nucleotides has been reported to occur rather normally in i-motif structures and to have a stabilizing effect on the motif. In fact, the first i-motif structure described by Gehring, Leroy and Guéron consisted of a stack of five $C:C^+$ base pairs capped at both ends by T:T base pairs.[56] The special interest behind T:T base pairs relies in the fact that they are isomorphic to the $C:C^+$ base pair, potentially acting as an extension of the core of the structure, not only at the ends of the $C:C^+$ stack, but also, although

more oddly, in the middle of it.[83] Other non-canonical base pairs such as Hoogsten A:T,[84] or A:A[85] and G:G[86] homo-base pairs, have been also reported to efficiently cap i-motif structures.

Of high interest for the stabilization of i-motif structures by capping interactions is the case of minor groove tetrads (MGT). The formation of these features can be understood as the association of two base pairs through their minor groove side within the i-motif structure. The first reported MGT was from the dimeric i-motif formed by the human B-box centromeric sequence d(TCCCGTTTCCA). The topology of this i-motif renders a head-to-head orientation, with a core stack of five C:C$^+$ base pairs and a slipped G:T:G:T MGT between the two lateral loops (see *Figure* 1-9).[86] The association of the two G:T base pairs occurs by the interaction of the guanidine residues through hydrogen bonding of their amino group and their N3 imino nitrogen. From that first description of MGT as capping and stabilizing agents of i-motif structures, it has been proven that their formation is compatible with other nucleobase combinations different than G:T:G:T.[87,88] The study and use of MGT as a stabilizing feature in i-motif structures has been a huge topic for research in our group, affording very promising results that will be introduced and properly discussed in following sections of this thesis.

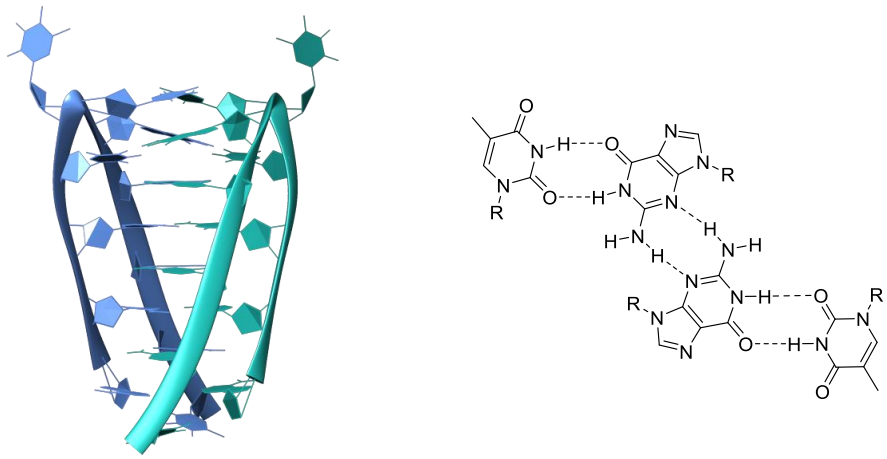


Figure 1-9. Three-dimensional structure of the centromeric i-motif dimer (left, PDB code 1C11). G:T:G:T minor groove tetrad (right).

1.1.5.2.1.C. Environmental factors

Besides intrinsic sequential factors affecting the stability of i-motif structures, externals factors such as the presence of cations and the ionic strength of the medium, the use of crowding agents or superhelical stress, have been reported to modulate the formation of i-motif structures.

An interesting approach for the stabilization of i-motif structures would be to test if small metal ions are able to fit between two cytidine residues, mimicking the hemiprotonated C:C$^+$ base pair by replacing the

H^+ atom. To this date, it has been reported by Waller's group that Ag^+ and Cu^+ cations are compatible with the formation of i-motifs at physiological pH and that, interestingly, the redox equilibrium between Cu^+ and Cu^{2+} can be successfully used to regulate the formation of the intercalated structure.[89,90]

On the other hand, if they are not directly taking place in the formation of the structure, cations do not have an influence, neither structurally nor stability-wise, on the i-motif structure because of its nature but due to their concentration in solution. Different studies observed a significant decrease in stability of i-motif structures from different telomeric and promoter sequences when increasing the ionic strength of the medium at pH values above the pK_a of cytidine.[68,80]

In another attempt to achieve the formation of i-motif structures near to physiological conditions, reproduction of the crowded cell environment *in vitro* has been proven to favor the formation of i-motif and G-quadruplex structures against B-DNA and to enhance its thermal stability at acidic pH.[91] Commonly used molecular crowding agents are high molecular mass polyethylene glycols (PEGs) but so far, their stabilizing properties have been reported to be insufficient by itself in order to afford stable i-motif structures at neutral pH.[92] However, more recent results showed formation of i-motif structures from promoter region sequences at neutral pH in presence of 40% w/w PEG_{200} and 1 mM $[Mg^{2+}]$.[93] Interestingly, single-walled carbon nanotubes (SWNTs) have been reported to be able to assist and stabilize $C{:}C^+$ formation, favoring the formation of i-motif structures in a molecular crowded solution (PEG_{1000}) at pH 7.0.[94] Furthermore, SWNTs have been proposed to act as effective regulation agents of telomerase activity both *in vitro* and *in vivo* by stabilizing i-motif formation.[95]

As it has been mentioned previously, the formation of non-canonical structures is highly favored under superhelical stress conditions which induce the DNA double helix to transitory adopt its single-stranded form. Unwound regions are then optimal to fold into non-B DNA structures. The reported experiments to demonstrate this theory *in vitro* consist on placing the i-motif-forming sequence in a supercoiled plasmid and to follow its formation under physiological conditions via chemical and enzymatic footprinting.[39] Results strongly support the initial hypothesis and show the simultaneous formation of G-quadruplex and i-motif structures from complementary MYC promoter region unwound strands.

1.1.5.2.2. Chemical modifications

The discrete introduction of chemical modifications in i-motif structures allows to specifically study how the alteration of different structural features affects the formation and stability of the structure. To this aim, a wide spectrum of modifications has been studied to potentially enhance i-motif stability. From natural-occurring alternatives like RNA i-motif or nucleobase lesions to more specifically directed modifications such as 2'-fluoro-arabino sugar rings have been tested.

A straightforward strategy to improve the stability of i-motif structures is to modify the chemical structures of the nucleobases directly involved in the formation of the motif. To this end, a recent study has evaluated how common natural base lesions affect the stability of the telomeric $d(CCCTAA)_3CCCT$ i-

motif.[96] Apurinic residues (apA), 8-oxoadenine (8oxoA) and 5-hydroxymethyluracil (5hmU) were introduced in the TAA loops, that act as stabilizing capping interactions. The results from this approach showed variable effect on the stability of the structure depending on the position where the lesion was introduced. Whereas thymine substitution with 5hmU does not significantly influence the thermal stability of the structure, structures containing apA and 8oxoA present T_m variations from -1.7 to +5.0ºC. The same study reports that, remarkably, simulating the naturally occurring replacement of cytidine residues from the C:C$^+$ stack with uracil, leads to a significant destabilization of the structure. The introduction of uracil in the core of the C:C$^+$ stack is so critical that, depending of the position where the replacement of cytosine residues is done, the decrease in T_m values is around 30ºC.

A widely used cytosine derivative in i-motif structures is 5-methylcitosine (5^mC), which is found to be present in C-rich sequences near transcriptional starting sites as an epigenetic modification. Different studies introducing different number of 5^mC in different intramolecular i-motifs, all led to stable i-motif formation under slightly acidic conditions with increased thermal stability and pH_T.[97] The N3 nitrogen atom proton affinity of 5^mC has been theoretically reported to be higher than the natural cytosine, due to the electronic enrichment of the pyrimidine ring provoked by the methyl group.[98,99] On the other hand, another commonly found epigenetic modification found in C-rich sequences, 5-hydroxymethylcitosine, affords less stable structures.[97] Other modifications introduced in the 5 position of the pyridine ring of cytosine include halogenated derivatives, which influence was studied in terms of BPEs of the respective heterodimers. The conclusion obtained for this theoretical study is that any halogenation in the 5 position of cytosine renders slightly weaker base pairs than the natural C:C$^+$, probably due to the electronegative nature of the halogen atoms.[99]

A more specific strategy for the stabilization of i-motif structures using chemically modified nucleobases, is based on introducing protonated cytidine analogues that would potentially mimic the formation of C:C$^+$ base pairs without protonation requirement. This approach is extensively discussed in the first chapter of this thesis and thus, a proper background will be provided later.

A recent report shows a phenoxazine analogue derivatized with a C8-aminopropanol tether.[100] The interest behind this derivative, named i-clamp, relies on its capacity to, apart from effectively form a stable base pair with a protonated cytosine residue, generate favorable electrostatic interactions through its aminopropanol clamp with the phosphate backbone of the opposite strand. Also, i-clamp – i-clamp hemiprotonated base pairs are suggested to be stably formed. The stability of intramolecular i-motif structures containing i-clamp residues is significantly position dependent, with i-motif formation under physiological conditions when i-clamp residues are introduced close to the 5'-end of the C:C$^+$ stack.

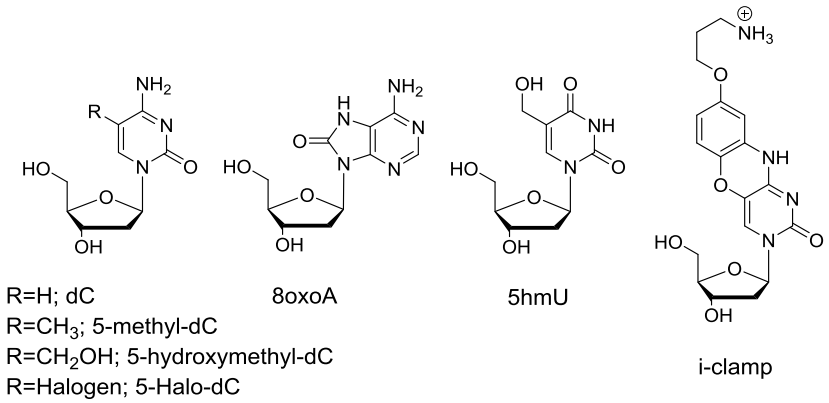


Figure 1-10. Chemical structure of some of the nucleobase modifications studied in the context of i-motif structures.

Another interesting strategy for introducing chemical modifications in i-motif structures is to modify the ribose rings of the sugar-phosphate backbone. The theoretical approach behind this strategy is to enhance the stabilizing interactions that already occur in natural DNA between sugar residues from different strands in the narrow groove. These stabilizing sugar-sugar contacts are of hydrogen bond nature and occur both inter- and intra-strand.[65]

More than strictly studying the effect on the stability of i-motif structure after introducing a hydroxyl group in the 2' position of the ribose ring, the feasibility evaluation of RNA i-motif formation is, obviously, of high interest for biology. Different studies point towards the same general conclusion: both full RNA sequences and partial ribose-substituted sequences form less stable i-motif structures than those of DNA i-motif. More specifically, reported data quantifies the loss in stability in RNA analogues compared to DNA i-motif structures. The melting temperature of d(CCCTCCCTTTTCCCTCCC) is almost 30ºC higher than its RNA analogue r(CCCUCCCUUUUCCCUCCC).[101] Moreover, the tetrameric i-motif formed by r(UCCCCC), which T_m at pH 4.3 is more than 50ºC lower than its DNA analogue, presents as major species the 5'-E topology, where the sugar-sugar interactions through the narrow groove are minimized.[102] Discrete substitutions of one or two DNA residues for ribonucleotides in d(TCCCCC) afforded less stable i-motifs (decrease of 5-20ºC) with maximum destabilization when the substitution was introduced in positions involved in back-to-back steps.[101]

Thus, although the C3'-*endo* sugar puckering commonly found in i-motif structures should favor the stable formation of RNA i-motif, the interactions of the 2'-hydroxyl groups through the narrow grooves of the structure lead to a notable destabilization of the motif. Even when introducing LNA in d(TCCCCC), an RNA derivative with a methylene bridge connecting the 2'-O oxygen and C4' carbon atom that locks the C3'-*endo* conformation of the sugar, the stability of the resulting tetrameric motif depends on the position in which the LNA residue is introduced,[103] suggesting that not only the sugar puckering but

more importantly the nature and the relative position of the 2' substituent, are crucial factors for the stability of the structure.

Interestingly, studies performed with 2'-fluororibose substitutions rendered more stable or slightly destabilized structures,[104] evidencing that the destabilization provoked by the 2'-OH substitutions is due to potential hydrogen bond interactions of the hydroxyl group and not due to hindrance factors. However, larger substituents do sterically disrupt the formation of the structure. This information is supported by the fact that when introducing 2'-arabino nucleotides in i-motif structures, the 2'-OH group of these residues is placed towards the major groove of the structure, resulting in greater stabilities than those observed from ribose-substituted motifs.[105]

Remarkably, Damha's group designed the so called 2'F-araC (2'-deoxy-2'-fluoroarabinocytidine) nucleotide based on the previous information.[106] The introduction of 2'F-araC in different positions of different inter- and intramolecular i-motifs rendered more stable structures than their unmodified analogues over a wide range of pH (up to pH 7 for the telomeric i-motif sequence). The enhanced stability of these structures was explained after NMR structural elucidation of the 2'F-araC-substituted motifs, that revealed that the 2'F-araC adopts a C2'-*endo* sugar conformation, placing the fluorine residue towards the major groove, not only avoiding repulsive interactions within the narrow groove, but also stabilizing the i-motif structure via favorable electrostatic interactions with the H2' hydrogen atom involved in sequential and inter-strand stabilizing contacts. Following the same approach, the same group and other designed different arabino-like derivatives that showed not significantly better results than the 2'F-araC residue.[107,108]

Since the intrinsic nature of i-motif structures affords remarkably narrow minor grooves, the repulsive electrostatic interactions between inter-strand phosphate groups is critical for the stability of the structure. Thus, a logical theoretical approach towards enhancing the stability of i-motif structures would be to introduce chemical modifications in the phosphate backbone in order to reduce these destabilizing interactions. In order to explore the scope of this hypothesis, a first set of modifications was proposed by Mergny and Lacroix.[109] They compared the feasibility and stability of different intramolecular i-motif structures containing phosphorothioate (S), methylphosphonate (M) and peptide (P) linkages with those of their natural phosphodiester analogues. Results from UV melting experiments revealed that only phosphorothioate bonds assisted i-motif formation with similar T_m values at pH ~ 6.5 than their natural analogues. However, methylphosphonate and peptide linkages did not fold into i-motif structures even at acidic pH. In the case of methylphosphonate, although the loss of the negative charge should alleviate the repulsion between adjacent narrow groove strands, the steric hindrance provoked by the methyl groups results in a significant destabilization of the structure.[109] Furthermore, more recent studies performed with peptide nucleic acids (PNA), concluded that the PNA analogue of the sequence d(TCCCCC), folds into a stable tetrameric i-motif at pH 4.1 - 4.5 while its unmodified analogue is able to form the motif up to pH 6.5.[110] This result is consistent with the fact that fully PNA-substituted

oligonucleotides destabilize or impede the formation of i-motif structures. Interestingly, a 1:1 DNA-PNA mixture has been reported to afford a PNA-DNA hybrid i-motif stabilized by the intercalation of two parallel heteroduplexes. This PNA-DNA hybrid i-motif can be observed in a pH range in-between that of its fully PNA and DNA analogues (pH 4.2 – 5.7).[111]

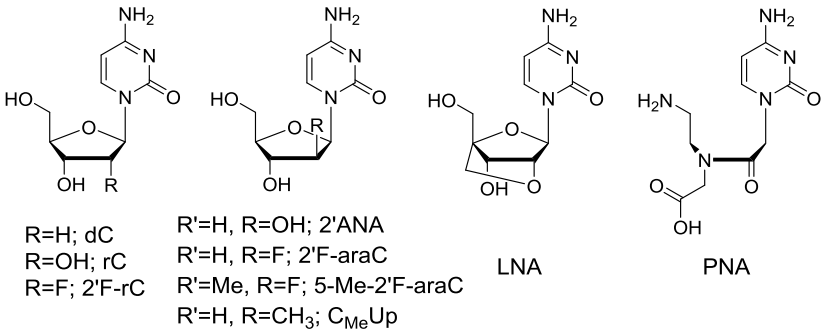


Figure 1-11. Chemical structure of some of the sugar modifications studied in the context of i-motif structures.

1.1.5.2.3. Nanotechnological applications

I-motif structures are very attractive from the nanotechnological point of view since their formation can be achieved, rapidly and reversibly, by local modulations of pH. In this respect, several nanodevices have been successfully designed based on i-motif-forming sequences.[112] Of special interest are those applications focused on utilizing i-motifs in therapeutics, mainly as specialized drug carriers and deliverers. In this section, an outline of the highlights reported in this research field will be briefly summarized.

The general approach for the design and construction of i-motif-based nanomachines, is to strategically link C-rich oligonucleotide sequences to different scaffolds in order to chemically manipulate (generally by pH modulation) the activity of the resultant material. The nature of the scaffolds utilized to fix the C-rich strands serves as criteria to classify the different nanostructures and to understand their applications.

A wide range of applications is focused on the design of smart materials and hydrogels. As examples, the aggregation via intermolecular i-motif formation of C-rich-derivatized branched molecules,[113] single-stranded long DNA molecules,[114] SWNTs,[115] PEGs[116] or other functionalized polymers[117] controlled by acidification of the medium, resulted in very stable stimuli-sensitive materials.

Another common strategy in this field is to attach C-rich sequences to metal ions. These metal-organic constructions generally focus their activity on crystalline network alterations triggered by i-motif

formation/denaturation. Gold nanoparticles (AuNP) have been proven to be very useful in this field through Au-S linkage of oligonucleotides as they are widely used as powerful biotechnological tools for drug delivery and sensing and imaging applications.[118,119]

Fluorescent FRET pairs can be also strategically incorporated into i-motif-based nanostructures in order to create smart pH-sensitive sensors. Tracking this idea, cell-penetrating AuNP functionalized with oligonucleotides complementary to i-motif-forming sequences were used as intracellular nanoprobes. Ideally, when internalized into an acidic cell medium, the i-motif formation would bring together the FRET pair and produce a traceable signal.[120] This approach has been similarly reproduced in different constructions where the i-motif is linked to different scaffolds such as hairpin-forming sequences[121] or biotin/streptavidin pairs.[122]

If the conjugation of i-motif structures takes place between a C-rich sequence and moieties capable of selectively recognize other molecules or higher order structures, the number of applications scales up rapidly and complex switches can be designed as specific nanobiotechnological detectors. Telomeric i-motifs have been reported to effectively conjugate with thiazole orange (TO) to render a detector for DNA higher order structures with strong imaging potential.[123] The recognition of tumorous cells can also be achieved by conjugating an i-motif-forming sequence to other specific cell-recognition DNA aptamer sequences.[124] Moreover, a smart approach based on the conjugation of i-motif DNA-derived linkers to iron oxide nanoclusters, afforded a very useful tool for cancer cells imaging. Thanks to the magnetic resonance imaging (MRI) properties of iron oxide nanoclusters, when the assembly is shattered via i-motif formation in contact with cancerous acidic cellular medium, the smaller nanoparticles present enhanced MRI signal. Notably, this strategy has reported efficient diagnosis for early stage small hepatocellular carcinomas (HCC).[125]

The therapeutic potential of i-motif-based nanostructures has been explored in several studies. The mechanism of these therapeutic agents is based on the pH-sensitive formation/denaturation of i-motif structures. The general approach is to capture, following different strategies, a drug into an i-motif-based nanostructure in a way that i-motif formation, triggered by pH decrease, results in the site-specific release of the drug. Some examples are already described focusing on delivering the DNA intercalating drug doxorubicin (DOX). Microcapsules based on $CaCO_3$ have been successfully engineered to encapsulate DOX, in a way that the stabilization of the structure is achieved via DNA bridges containing i-motif-forming sequences hybridized into complementary duplexes. Under acidic conditions the structure collapses due to i-motif formation and frees DOX into the medium.[126] Similarly, AuNP clusters functionalized with both G-quadruplex and i-motif-forming sequences (Au-GIs) have been also designed to deliver DOX under acidic endosomal pH.[127] DOX can also be trapped into PEG-functionalized nanostructures that collapse due to i-motif formation.[128] These later examples have been proven to be effective *in vitro* and *in vivo* for the remission of different tumorous cells.

1.1.5.2.4. Biological implications

Contrary to the first theories assumed after its discovery, now is well stablished that the formation of i-motif structures can be achieved under physiological conditions. Moreover, since C-rich sequences are found in strategic sites of the genome as complementary to the better studied G-quadruplex forming sequences, the hypothesis that i-motif structures play an active and relevant role *in vivo* has gained many followers.

1.1.5.2.4.A. i-Motif structures in human cells

Two different recent studies have proposed two different approaches for the visualization of i-motif structures in human cells.

The first evidence of the visualization of stable i-motif structures in cellular environment was reported by Dzatko *et al.* using NMR spectroscopy.[63] The experiments that they performed consisted in the transfection of purposefully fluorescent marked DNA sequences from human promoter regions such as DAP, HIF-1α, or PDGF-A into HeLa cells. By using confocal microscopy, the researchers were able to confirm that the previously annealed structures remained unaltered when introduced in the cell and, moreover, that they were localized in the nuclei of the cell without producing any substantial alterations to neither the cellular medium nor the cell viability. Once transfected, the *in-cell* NMR experiments confirmed that the DAP, PDGF-A and JAZF1 sequences showed the distinguishing imino signals of i-motif structures corresponding to the formation of hemiprotonated $C:C^+$ base pairs up to 35ºC. On the contrary, for the sequence from the HIF-1α gene promoter, no i-motif formation could be confirmed using *in-cell* NMR at that temperature.[63]

The other approach for the visualization of i-motif structures in human cells applies a similar strategy than the previously mentioned report for the visualization of G-quadruplex structures.[46] The idea behind this experiment was to select from a wide library of compounds a highly selective antibody that exclusively binds i-motif structures. After ELISA+BLI characterization of the leading compound, which they named iMab, the authors report that the iMab affinity for other DNA secondary structures is very low or null and that this antibody can bind i-motif structures with different topologies.[64] Moreover, the affinity of iMab for i-motif structures is pH-dependent and does not alter, as tested by CD experiments, the overall structure of the motif. After the *in vitro* characterization of the i-motif-iMab interaction that rendered affinity constant values of ~0.5 – 10 nM, the indirect visualization by immunofluorescent staining displayed punctuate foci attributable to i-motif structures in the nuclei of different cell lines (*i.e*. MCF7, U2OS and HeLa). Interestingly, the treatment with both DNase I and RNase I suggests that, even if most of the foci disappeared with DNase I, some of the foci could correspond to RNA i-motifs.[64]. In the same study, the relation between the formation of i-motif structures and the cellular cycle stage is also explored. Whereas the G-quadruplex formation seems to be predominately achieved during S-phase, i-motif formation seems to occur mostly during G1/S (late G1) phase. That means that i-motif

structures would play their role predominantly during transcriptional processes whereas G-quadruplexes would be more involved in DNA replication.[46,64]

1.1.5.2.4.B. i-Motif localization and function in the human genome

The exploration of C-rich sequences throughout the human genome has been performed by different research groups, with the aim of both measure their prevalence and to obtain a first idea of their possible activity by connecting it with their location. Waller's group reported a screening using the consensus sequence $d(C_5X_{1-19})_3C_5$ introduced on the search tool Quadparser. The search afforded 5125 hits with a significant 12.4% corresponding to promoter regions.[72] A different search was performed by Burrow's lab by using bioinformatics analysis focused on localizing C-tracts of different lengths (*i.e.* dC_{15-81}). The results of the analysis showed that 769 dC_n sequences exist in the human genome with high frequency on gene promoters and other untranslated regions.[74] *Table* 1-1 summarizes some of the most studied genomic sequences known to fold into i-motif structures. Out of all of them, the features and implications of the most studied ones, including not only sequences from promoter regions but also telomeric and centromeric sequences, will be highlighted in the next lines.

NAME	SEQUENCE
hTelo	$d(TAAC_3TAAC_3TAAC_3TAAC_3)$
A-Box	$d(TTCCT_4CC(T)ACCATAG)$
CENPB-Box	d(TCCCGTTTCCAACGAAG)
Dodeca	$d(C_3GTACTGGT/C_3GTACTCGGT)$
BCL-2	$d(CAGC_4GCTC_3GC_5TTCCTC_3GCGC_3GC_4T)$
MYC	$d(TC_4ACCTTC_4AC_3TC_4AC_3TC_4A)$
PDGFR-β	$d(GCGTCCAC_3TC_3TGC_4GCCGC_7TTCTC_3AGC)$
KRAS	$d(GC_3GGC_5GCTCCTC_5GCCGGC_3GGC_3GGC_5TCCTTCTC_4G)$
HRAS	$d(CGC_3GTGC_3TGCGC_3GCAAC_3GA)$
VEGF	$d(GAC_4GC_5GGC_3GC_4GG)$
PDGF-A	$d(CCGCGC_4TC_5GC_5GC_5GC_{13})$
HIF-α	$d(CGCGCTC_3GC_5TCTC_4TC_4GCGC)$
JAZF1	$d(C_8GC_5GC_5GC_3TC_6)$
DAP	$d(C_5GC_5GC_5GC_5GC_5)$
ILPR	$d(TGTC_4ACAC_4TGTC_4ACAC_4TGT)$
KIT	$d(C_3TCCTC_3AGCGC_3AC_3T)$
RET	$d(CCGC_5GC_4GC_4GC_4TA)$
Rb	$d(GCCGC_3A_4C_6G)$
RAD17	$d(CCAC_9GC_9GGA)$

Table 1-1. Most studied i-motif-forming sequences found in the human genome.

As it was previously discussed for the case of G-quadruplex structures, telomeres are crucial for the integrity and lifetime of the chromosomes during different cell processes.[43] Thus, complementary to the tandem repeat d(TTAGGG), capable to fold into a G-quadruplex motif which stabilization has been proven to inhibit telomerase activity, the telomeric sequence d(CCCTAA) is reported to analogously form an i-motif structure.[68] Different versions of the telomeric hTELO i-motif forming sequence have been studied *in vitro* depending on the number of residues selected for the study but, independently of it, the motif consists on the intercalation of four CCC tracts connected with three TAA loops affording a six-step C:C$^+$ stack.[60] The structure shows remarkable thermal stability at pH values close to neutrality.[72] Although the targeting of the telomeric i-motif with ligands for effectively stabilize its formation has not been studied as thoroughly as in the case of G-quadruplexes, carboxyl-modified SWNTs have been reported to be the first molecules to selectively bind and stabilize the hTELO i-motif structure.[95] Carboxyl-modified SWNTs also favor the formation of G-quadruplex structures by enabling duplex unwinding. The simultaneous formation of G-quadruplex and i-motif structures captured by SWNTs impedes telomere elongation by structural disruption which subsequently provokes cell death.[129] More research is needed to fully understand the biological implication of hTELO i-motif specially for considering it as an effective therapeutic target against proliferation of cancer cells.

The architecture of centromeric chromatin has been recently suggested to be based on non-canonical DNA structures.[130] In this respect, the formation of i-motif structures might be directly involved in the constitution of the multiprotein complex. The centromeres, although notably varying in primary structure between different species, are proposed to play a very important role during chromosome segregation and posterior assembly during cell division processes. In human centromeres, the so-called alpha satellite is an AT rich tandem repeat of 171 base pairs which contains a GC rich segment of 17 base pairs. A box and B box (also CENP-B box, for being the binding site of the CENP-B centromeric protein) are two different variants of this segment. At the same time, the A box presents two different versions differing only in one nucleotide. Remarkably, CENP-B box evolved from the A box and is absent in human Y chromosome.[131] The structure of the C-rich strands of both A and B boxes has been elucidated by NMR studies and afforded dimeric i-motif structures.[86,132] While the A-box dimeric i-motifs adopt a head-to-tail orientation, that is with the loops at opposite sides of the structure, stabilized by capping A:T base pairs,[132] the B box i-motif prefers a head-to-head orientation stabilized by a G:T:G:T minor groove tetrad.[58,86] Very similar to these i-motif topologies observed in human centromeric sequences, the centromeric region of chromosome 3 of *Drosophila melanogaster* contains a tandem repeat named dodeca after it is composed of 11/12 base pairs repeat. Both 11-mer and 12-mer fragments have been reported to fold into dimeric i-motif similarly to the A and B human boxes.[133] Altogether, the common observations found for different centromeric sequences give proof for a potential structural function of i-motif structures. This hypothesis suggests that i-motif formation would serve as a long-range interaction between associated nucleosomes at the ends of the centromeres.[132]

Due to the relatively high prevalence of C-rich sequences close to gene promoter regions, it is reasonable to believe that i-motif structures major role *in vivo* has to do with the regulation of gene expression as a structural recognition unit.

Probably, the most studied promoter sequence that folds into an i-motif structure is the one belonging to the P1 promoter region of the B-cell lymphoma-2 (BCL-2) oncogene. This oncogene overexpression is found in several cancer cells provoking apoptotic resistance. On the contrary, it is found under-expressed in some diseases of the nervous system such as Alzheimer´s or Parkinson's. The i-motif structure described for the BCL-2 promoter sequence consists on an intramolecular motif with seven $C:C^+$ base pairs in its core and *class II* (8:5:7) loop topology.[61] Interestingly, the same sequence can adopt a hairpin-like alternative structure. Hurley's group has performed extensive studies focused on how the equilibrium between the two secondary structures could be modulated and which repercussion would this have on the transcriptional regulation.[134] From the screening of close to two thousand small molecules, they isolated two compounds, IMC-48 and IMC-76 that interact and favor the formation of the i-motif and hairpin structures, respectively. Moreover, the authors were able to test the effect on the transcription of the BCL-2 promoter of these two compounds in lymphoma cell lines. The results conclude that IMC-48 induces transcriptional activity by hypothetically locking the i-motif structure, whilst IMC-76 favors the formation of the alternative hairpin which suppresses BCL-2 transcription.[134] In a parallel study also performed by Hurley's group, the heterogenous ribonucleoprotein LL (hnRNP LL), that functions as a transcriptional activator factor, was reported to selectively bind the i-motif structure formed by the BCL-2 promoter.[135] Two of the four RNA recognition motifs (RRM) of hnRNP LL were described to bind the two lateral loops of the BCL-2 i-motif resulting, as tested by CD and bromine footprinting experiments, in the unfolding of the i-motif structure. Thus, as a summary, IMC-48 shifts the equilibrium of the BCL-2 promoter sequence to the i-motif structure, which can be specifically recognized and unfolded by hnRNP LL that ultimately activates gene expression. These consecutive studies give example for the most detailed mechanism of regulation of transcriptional activity in which i-motif structures play a crucial role.[134,135]

Another well-studied promoter sequence that potentially folds into an i-motif structure is that of the nuclease hypersensitive element (NHE) III_1 region from the MYC (from the Myelocytomatosis virus) oncogene. MYC functions are particularly focused on cell growth and cellular cycle maintenance and can be found overexpressed in several cancer types. NHE III_1 is of special interest among the other nuclease hypersensitive elements since regulate around 90% of MYC transcription and contains a GC-rich fragment that adopt, potentially, G-quadruplex and i-motif structures. While the G-quadruplex binding by nucleolin turns off gene expression, in an analogous way to hnRNP LL with BCL-2 i-motif, hnRNP K activates transcription. The 4CT C-rich element from NHE III_1 folds into an i-motif structure with eight $C:C^+$ base pairs core and *class II* loop topology (5:5:5).[136] Hurley's group proposed that the mechanism for the activation of MYC expression starts with the unwinding of the duplex instigated by the binding of SP1 which apparently provokes superhelical stress to the double helix. Once the i-motif

structure is formed, hnRNP K recognizes it through CCCT runs. Interestingly, the 4CT i-motif-hnRNP K complex is not strong enough to unfold the i-motif and the assistance of a third CCCT run from the 5CT element must be added to the complex for successfully activate transcription.[136]

Apart from these two most studied promoter sequences in which the recognition of an i-motif structure induces gene expression, an increasing number of examples suggest the same regulatory function for i-motif/G-quadruplex structures as ON/OFF switches for transcription activation for other promoter sequences. Similarly to MYC, a NHE element from the promoter of the platelet-derived growth factor receptor β (PDGFR-β) has been reported to fold into an i-motif structure that contains, like MYC, CCCT segments that can be recognized by hnRNP K.[137] Overexpression of PDGFR-β is commonly related to vascular, fibrotic and tumorous pathologies. The i-motif from the C-rich strand of the PDGFR-β promoter region has been characterized as an intramolecular motif with six $C:C^+$ intercalated base pairs and *class II* topology (5:10:8) using T-to-C mutations to avoid multiple topologies observed from the wild type sequence. The authors of this study report an ellipticine derivative (NSC309874) capable of downregulating gene expression of PDGFR-β by binding the i-motif structure from the NHE.[137]

RAS gene family codes for small GTPase proteins responsible for transmitting cellular signals. The overexpression of RAS implies activation of other genes involved in proliferation, differentiation and cell endurance that ultimately lead potentially to cancer disease. KRAS and HRAS are different subfamilies of RAS. They are located in different chromosomes in humans and have been suggested to be related to different types of cancer such as lung or bladder cancers.[138] Within the KRAS promoter region, three different C-rich regions can be observed, being the mid one the most stable one. Moreover, the i-motif formed by this KRAS promoter mid region is reported to be in a dynamic equilibrium with alternative hairpin and hairpin-i-motif hybrid structures controlled by pH.[139] A similar equilibrium has been described for the i-motif formed in the promoter region of HRAS.[140] The proposed topologies for KRAS and HRAS are significantly different. While KRAS promoter mid region is proposed to fold into an intramolecular i-motif with eight $C:C^+$ base pairs and long loops topology (7:16:7), HRAS promoter forms an i-motif structure with six intercalated $C:C^+$ base pairs and shorter loops (3:4:4).[139,140] KRAS interactions with hnRNP K have been accurately studied and the small molecule nitidine has been proposed as a transcriptional regulator of KRAS by binding the i-motif/hairpin hybrid structure of its mid region promoter. Notably, a sequential mechanism involving the three C-rich regions of the KRAS promoter was proposed somehow analogous to MYC expression. The hypothesis, based on the stability of the different i-motif structures from the different regions, is that near and far regions regulate the access of hnRNP K to the transcriptional active mid region.[139] As for HRAS, it has been reported that the nuclear factor hnRNP A1 is fundamental for its transcription and that the mechanism to trigger HRAS expression is similar to the previously described gene promoters, where hnRNP proteins not only recognize the i-motif structure but provoke its unfolding by binding to their lateral loops.[140]

The human vascular endothelial growth factor (VEGF) has been found overexpressed in tumorous tissue and metastatic cancers. GC-rich fragments capable to fold into quadruplex structures have been described upstream the transcription start site of VEGF. Thus, another example of i-motif structure in the promoter region of an oncogene responsible for the regulation of its transcription was reported. The intramolecular i-motif from the VEGF promoter region presents six intercalated $C:C^+$ base pairs and *class I* loop topology (2:3:2). As in other studies, hnRNP K was found to regulate VEGF transcription by selectively binding the i-motif structure formed in its promoter region.[141]

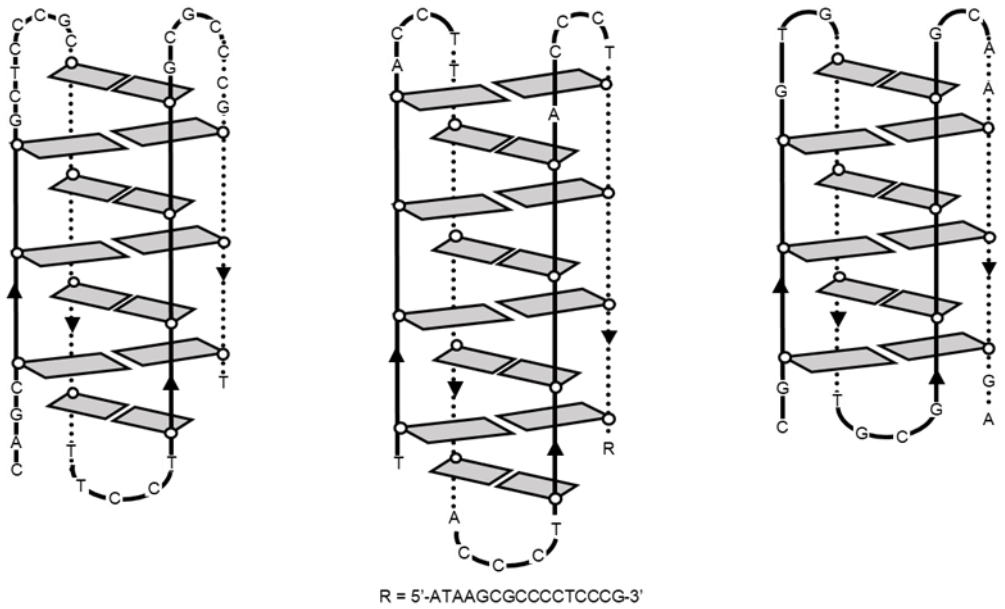


Figure 1-12. Schematic representation of the predicted i-motif structures of some of the genes presented in this section: BCL-2 (left), MYC (center) and HRAS (right).

The above described examples show how the transcriptional mechanism of different genes is regulated by i-motif structures in their promoter regions and, more interestingly, how the interception of these structures with small molecules could lead to therapeutic applications. Moreover, many other examples of oncogene promoters that have not been extendedly discussed in this introduction have been reported recently to also fold into i-motif structures. Altogether, tangible evidences of the role of i-motif structures *in vivo* have been reported recently settling the basis for giving credit to i-motif structures as a crucial biologically relevant nucleic acid structure.

1.1.5.2.5. i-Motif targeting with small molecules

The design of small molecules for selectively binding i-motif structures is an important breakthrough in the field in order to potentially consider i-motifs as therapeutic targets. The structural and topological features of i-motif structures allow the differentiation of four specific kinds of interactions with small molecules. Theoretically, depending on the nature of the molecule, the binding to the i-motif structure might occur through the grooves of the motif, by stacking interactions at the ends of the stacks, by intercalation or through the loops of the structure.[142]

In this respect, the previously mentioned compound IMC-48 capable of stabilizing the formation of the BCL-2 i-motif and, by doing it, activate gene transcription, shows predilection for the central loop of the structure.[134] On the other hand, PB1, another reported compound to bind the BCL-2 i-motif seems to interact with the i-motif through its major groove.[143] Moreover, a tetraoxazole derivative has been recently reported to bind the BCL-2 i-motif preferentially through its lateral loops.[144]

Apart from the previously mentioned NSC309874 molecule that binds the i-motif formed by the PDGFR-β promoter,[137] or nitidine that is able to regulate KRAS expression,[139] other examples have been reported. B19, an acridone derivative, showed regulation properties by binding the MYC promoter i-motif.[145] Also, mitoxantrone is reported to selectively interact with the MYC i-motif over other secondary structures *in vitro*[146] and finally, a screening of heterocyclic compounds for selectively binding the HRAS i-motif afforded some interesting hits.[147] Compared to G-quadruplexes, i-motif structures present less extended planar π systems that, theoretically, would restrict potential stacking interactions with small molecules. Furthermore, the short distance between $C:C^+$ base pairs is a handicap for the intercalation of ligands. However, planar compounds like porphyrins[148] or dyes like thiazole orange[149] are suggested to bind i-motif structures through stacking or intercalating interactions.

Finally, well-known G-quadruplex binders like berberine, BRACO-19, mitoxantrone, PDS or TMPyP4 were tested against i-motif structures. Even if the results of this study reveal that these compounds have the capacity to interact with i-motifs at neutral pH, all the interactions led to the destabilization or unfolding of the motif.[150,151]

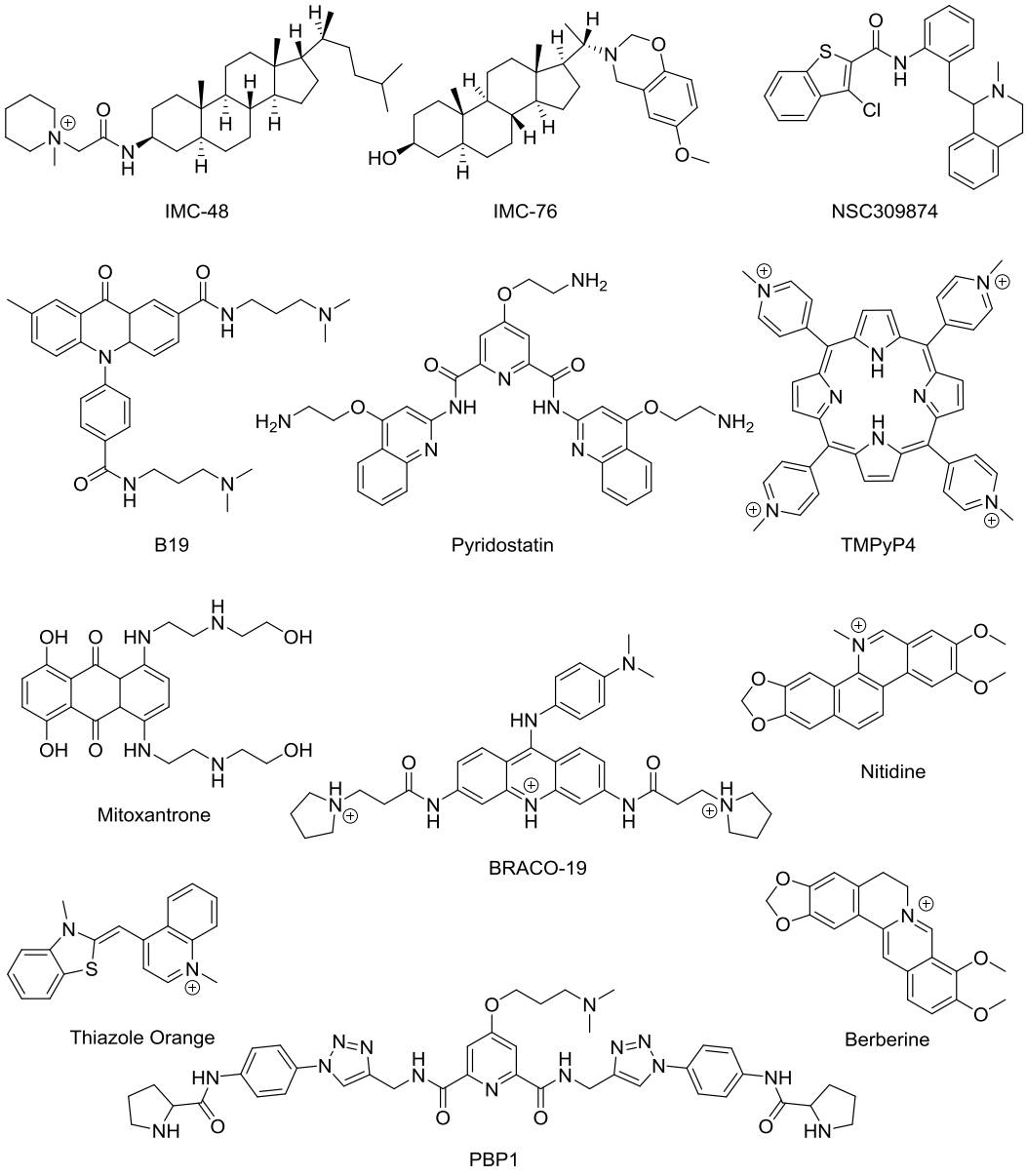


Figure 1-13. Different i-motif binding molecules.

1.2. OBJECTIVES

Given the increasing interest in i-motif structures, obtaining such structures as well as detailed structural information under physiological conditions have become hot topics in the structural biology field.

In this context, the main objectives of this thesis are focused on the design and detailed characterization of several oligonucleotide sequences that may form stable i-motif structures at neutral pH. The starting point is the mini i-motif structure, extensively studied in the research group, that exhibits unusual high stability versus temperature and pH. With the aim of getting deeper insights in this type of structures and enhance their stability at neutral pH conditions, two well-differentiated approaches are followed.

The first approach relies on the use of cytidine analogues. A first strategy consisted in the incorporation of a neutral analogue of protonated cytidine (pseudoisocytidine, psC) occupying specific positions of the motif. The 3*H*-tautomer of psC, thanks to the extra hydrogen-bond donor, can form neutral base pairs stabilized by three hydrogen bonds yielding a base pair completely isomorphic to hemiprotonated C:C$^+$. The aim is to explore whether the formation of pH-independent i-motif structures is possible by incorporating neutral base pairs. On the other hand, a fluorescent cytidine analogue (1,3-diaza-2-oxophenoxazine, tC^o), capable to hybridize as a cytosine and maintaining the ability to form Watson-Crick as well as hemiprotonated base pairs, will be tested as an internal probe for the characterization of local environments within i-motif structures. As a more ambitious goal, the use of tC^o for the visualization of these structures in cellular media will be explored.

The second approach focuses on studying the compatibility and stabilizing effect of different types of minor groove tetrads as capping elements in i-motif structures. To date, minor groove G:T:G:T tetrads have been observed in an i-motif structural context but a number of different minor groove tetrads have been observed in other non-canonical structures. In this context, both linear and cyclic oligonucleotides will be prepared for evaluating different nucleobase associations in the tetrads. Moreover, longer repeat sequences will be studied in order to explore if the formation of stable unimolecular mini i-motifs containing such tetrads is feasible.

2. FUNDAMENTALS

2.1. STRUCTURE OF NUCLEIC ACIDS

The aim of this section is to highlight several structural concepts that are important to understand the structural versatility of DNA. From very basic concepts to deeper structural insights, that are common and useful throughout all the chapters of this thesis, are detailed in the following sections.

2.1.1. PRIMARY STRUCTURE[152]

Nucleic acids can be structurally understood as biopolymers formed by the repetitive association of nucleotides which are, at the same time, composed by three different molecular fragments: a sugar, a heterocyclic aromatic base (commonly named nitrogenous base or nucleobase) and a phosphate group. β-D-ribose is characteristic for RNA strands whereas DNA molecules only contain β-D-2'-deoxyribose. The sugar ring atoms are numbered as shown in *Figure* 2-1. For those positions with geminal hydrogens, a differentiation between X' (*pro-S* hydrogen) and X'' (*pro-R* hydrogen) is made. Five different canonical nitrogenous bases divided into two groups (depending on its chemical structure) can be incorporated in nucleic acids. These are either purine bases: adenine (A) and guanine (G) or pyrimidine bases: cytosine (C), thymine (T) and uracil (U). Thymine is only present in DNA molecules whereas uracil is specific to RNA. The numeration of the functional groups of each nucleobase is assigned according to that of the atom of the aromatic ring which is bond to (*Figure* 2-1). Oligonucleotide sequences are always written starting from the 5'-end to the 3'-end. The connectivity of the polymer works as it follows: the phosphate group links the 3'-hydroxyl group of one nucleotide with the 5'-hydroxyl group of the following nucleotide via a phosphodiester bond. Nucleobases are covalently bound to the C1' carbon of the sugar. Purine bases are linked through the N1 nitrogen and pyrimidine bases through the N9 nitrogen (*Figure* 2-2).

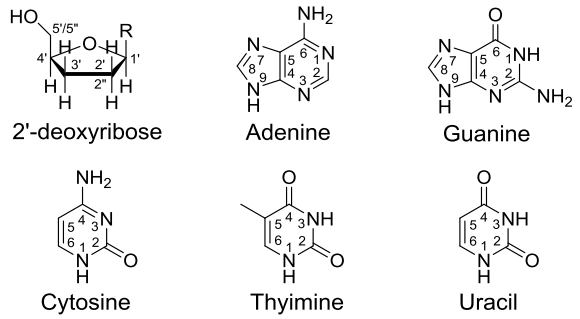


Figure 2-1. Nomenclature and numeration of 2'-deoxyribose and natural nucleobases.

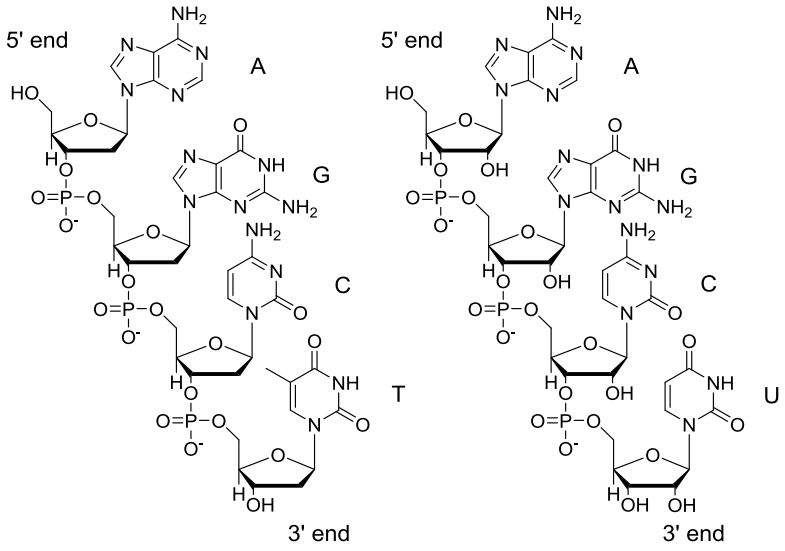


Figure 2-2. Representation of the connectivity of DNA (left) and RNA (right) strands.

2.1.1.1. CONFORMATIONAL FEATURES

2.1.1.1.1. Sugar conformations

The furanose ring can adopt multiple conformations. This characteristic, known as sugar puckering, is largely responsible for the great flexibility of DNA (and RNA) strands. The five atoms of the sugar ring are not coplanar, and two major conformations are adopted. The envelope (E) conformation maintains four atoms in the same plane and one out of it. On the other hand, the twist (T) conformation exhibits two atoms out of the plane and opposite to each other. Depending on whether the atoms out of the plane are on the same side of the plane than the C5' carbon or opposite of it, they are named *endo* (written as a superscript) or *exo* (written as subscript) respectively.

Pseudorotation angle (P) is a concept that combines the five endocyclic torsion angles ($\nu 0$, $\nu 1$, $\nu 2$, $\nu 3$ and $\nu 4$, represented in *Figure* 2-3) which are dependent of each other, and it is used to define the puckering of the sugar.[153] The phase of the pseudorotation angle, together with the puckering amplitude (ν_m), is used to describe the whole conformation of the sugar ring. Theoretically, the pseudorotation angle can take any value between 0º and 360º however it has been experimentally observed that two major conformations are preferred. These are north (N) or C3'-*endo* ($0<P<36º$) and south (S) or C2'-*endo* ($144º<P<190º$) (*Figure* 2-4, left and top right). The interconversion between the two conformations may occur in solution through less stable conformations and the preference for one of the conformations can be outlined. In general terms, purine bases prefer C2'-*endo* sugar conformations while pyrimidine bases tend to adopt C3'-*endo* conformations. The nature of the sugar also affects its conformation: the

C2'-*endo* conformation is commonly adopted in DNA structures whereas ribonucleotides favor C3'-*endo* conformations.

2.1.1.1.2. Glycosidic bond conformation

The conformation of the dihedral angle of the glycosidic bond (χ) is also important for the global structure of nucleic acids. For purine bases is defined by the atoms O4'-C1'-N9-C4 while for pyrimidine bases, by the atoms O4'-C1'-N1-C2. Although any value of χ is possible since the rotation has a relatively low energy barrier, two preferred relative conformations are found for nucleobases (*Figure* 2-4, bottom right). In the *syn* conformation (χ = 0±90º), the nitrogenous base eclipses (totally or partially) the furanose ring. In the *anti* conformation (χ = 180±90º) the sugar and the nucleobase are in opposite positions. Even though the *anti* conformation is more stable due to obvious steric reasons and thus, more present in different nucleic acid structures, the *syn* conformation is also found in various structures.

2.1.1.1.3. Sugar-phosphate backbone conformation

From the phosphorus atom in the 5'-end to the oxygen atom in the 3'-end, six dihedral torsion angles (α, β, γ, δ, ε and ζ) define the sugar-phosphate backbone conformation for one nucleotide. These angles are not independent of each other and thus, any conformational change in the overall secondary structure of the polynucleotide provokes a concerted change in the sugar-phosphate backbone.

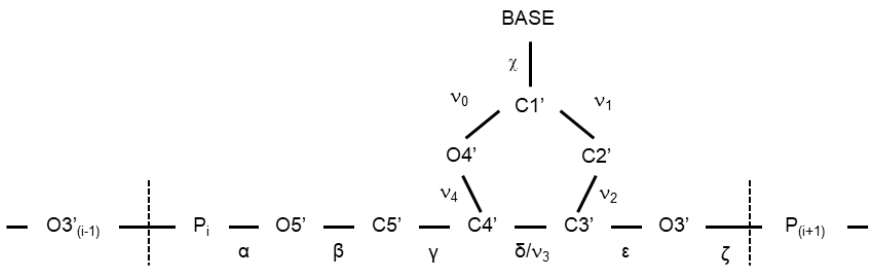


Figure 2-3. Torsion angles of sugar-phosphate backbone (α, β, γ, δ, ε, and ζ), glycosidic angle χ and endocyclic angles (ν_0, ν_1, ν_2, ν_3 and ν_4).

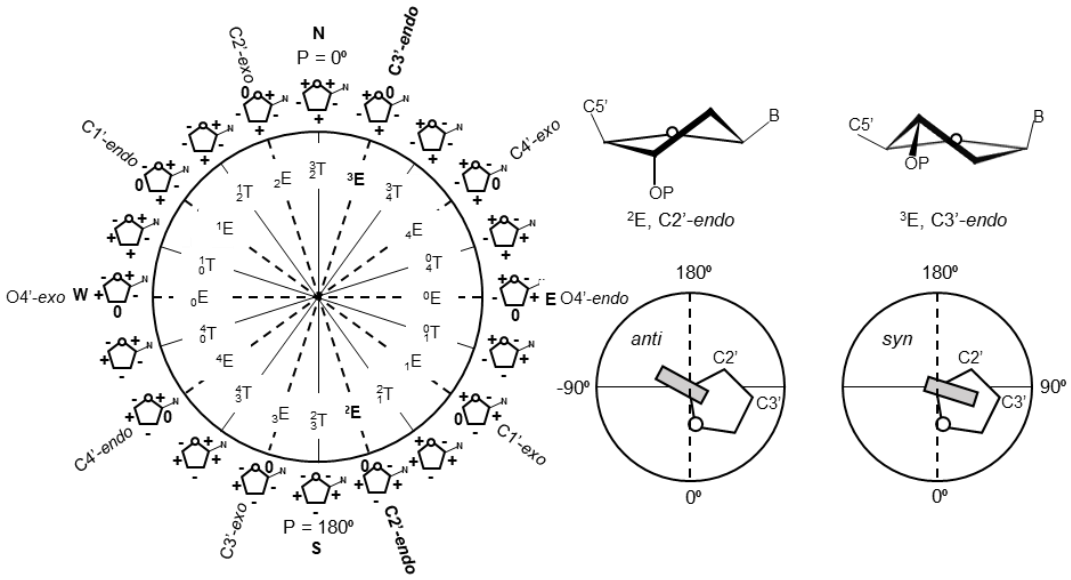


Figure 2-4. Pseudorotation cycle of the ribose ring (left), C2'-*endo* and C3'-*endo* ribose conformations (top right) and anti and syn conformations of the nucleobases (bottom right).

2.1.1.2. BASE INTERACTIONS

It is essential to discuss how nucleobases interact and recognize each other in order to understand how the different secondary structures of the nucleic acids are formed. It is the intrinsic chemical nature of the nitrogenous bases that gives them such a functional and structural relevance. Nucleobases association occurs through different hydrogen bonding patterns and their stacking interactions through London dispersion forces and hydrophobic effects.[154,155] These interactions are extendedly explained in the following sections.

2.1.1.2.1. Hydrogen Bond interactions

Hydrogen bond interactions are electrostatic and happen when a hydrogen atom is shared between two heteroatoms. Acceptor atoms must be highly electronegative and have a free non-bonding electron pair, whereas donor atoms are also electronegative atoms covalently bound to hydrogen. The association of the three atoms is geometrically linear and the distance between the heteroatoms oscillates between 2.5 and 3.5 Å. Energetically, hydrogen bonds are slightly stronger than Van de Waals interactions and significantly less stable than covalent bonds, thus, they are reversible interactions. The hydrogen bonds that are formed between the bases of nucleic acids involve hydrogen atoms from the amino and imino groups and oxygen and nitrogen atoms from the carbonyl and imino groups (N-H···O and N-H···N), respectively.[156]

Watson and Crick described the canonical (WC) base pairing pattern for G:C and A:T base pairs when they first described the DNA double helix structure.[1] Nevertheless, taking into account their acid-base properties and possible tautomeric forms, multiple different associations are possible between nucleobases: WC, reverse WC (rWC), Hoogsteen (H), reverse H (rH), *wobble* (W) and several *mismatches (Figure* 2-5 *and Figure* 2-6*)*.[157] WC G:C and A:T base pairs are isomorphic and lead very restrictedly to regular double helix with very well-defined major and minor grooves. Through these grooves, additional interactions may occur with other nitrogenous bases. On the other hand, purine bases can flip and form an alternative set of hydrogen bonds with the pyrimidines leading to Hoogsteen base pairs (*Figure* 2-6). A:T base pairs at neutral pH and G:C$^+$ base pairs under acidic conditions can adopt Hoogsteen patterns that have been observed in transitory conformations of DNA helical duplexes.[158] Reverse configurations of WC and Hoogsteen associations occur when one of the nucleobases is rotated 180º. In *wobble*-like base pairs, one of the nucleobases is displaced. Out of A:T and G:C base pairs, a great number of mismatched base pairs can also be found: G:A, G:T, G:G, T:T, C:C, T:C or A:C. This wide range of hydrogen bonding patterns is very relevant for the secondary structure of nucleic acids as they lead to some non-canonical structures. Of particular interest for the structures presented in this thesis, formation of G:T, G:G minor groove contacts and hemiprotonated C:C$^+$ base pairs will be extensively discussed throughout this manuscript (*Figure* 2-7).

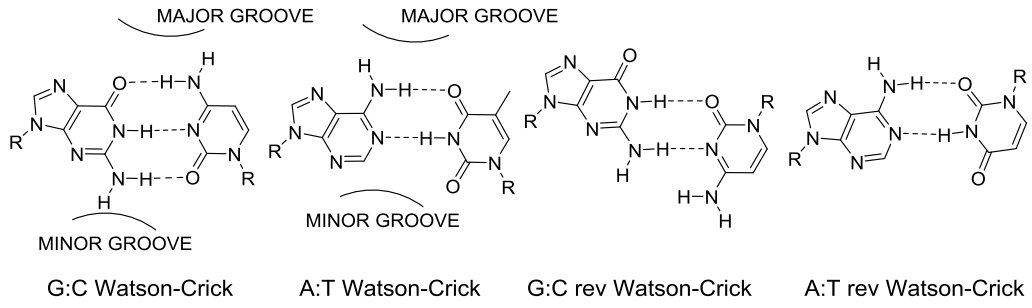


Figure 2-5. Watson-Crick and reverse Watson-Crick base pairs.

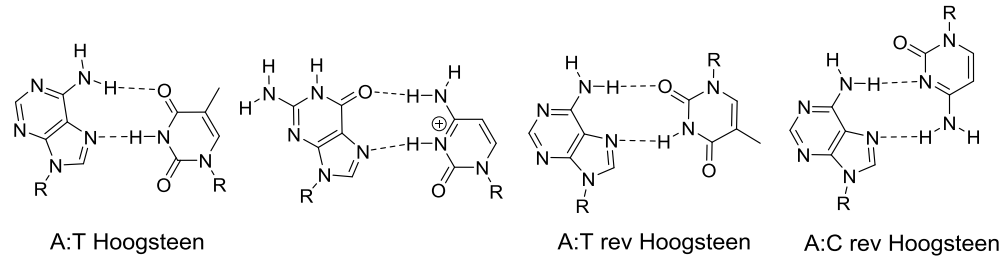
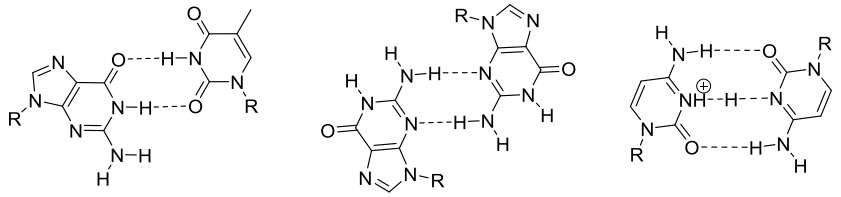


Figure 2-6. Examples of Hoogsteen, and reverse Hoogsteen base pairs.



G:T *wobble* G:G *mismatch* C:C$^+$

Figure 2-7. G:T, minor groove G:G and hemiprotonated C:C$^+$ base pairs.

2.1.1.2.2. Stacking interactions[155]

The aromatic nature of nucleobases allows them to interact through their π systems. The strong quadrupole moment together with the London dispersion forces are responsible for stabilizing this kind of interaction. Electrostatically, the π-π stacking occurs in a manner that the repulsive forces between nucleobases is minimized and thus, a geometric pattern is observed altogether with the one described by the hydrogen bonding. Stacking interactions may take place between bases of the same strand as well as between non-consecutive nucleobases. It is worth mention how the hydrophobic effect also contributes to this kind of interactions. In aqueous solution, nucleic acids fold in a manner that water is excluded from the core of the structure and surrounds the exterior of the biomolecule, generally in contact with the phosphate groups. This effect is favored by entropic factors and facilitates the stacking of the nucleobases. Energetically, purine bases exhibit stronger stacking interactions, suggesting that the extension of the π system is an important factor for the stability of the interaction. Overall, stacking interactions play an important role in the global structure of the nucleic acids.

2.2. CHARACTERIZATION TECHNIQUES FOR NUCLEIC ACIDS

A wide range of techniques have been used throughout this thesis with the aim to extract as much information as possible from the nucleic acids that were studied. Stability and structural data obtained from different methods were essential to understand the behavior of the structural motifs covered in this manuscript. This section attempts to summarize all the theoretical background of the different techniques that are transversal in this thesis. Experimental details are not described in this section.

2.2.1. UV SPECTROSCOPY[159]

The most significant chromophores that are present in nucleic acids are the nucleobases. Their aromaticity gives rise to intense electronic transitions between the bonding and non-bonding π orbitals (π → π*) with maximum absorbance bands at around 260 nm and molar extinction coefficients

magnitudes of about 10^4 ($M^{-1} \cdot cm^{-1}$). Thus, by measuring the absorption of a certain solution, one can perform a direct polynucleotide quantification obtaining experimental values of optical density ($OD_{260} = \varepsilon_{260} \cdot n$) (see *Section* 8.4.2 in Methods).

Another interesting utility of this electronic feature relies on the changes in the UV absorbance that are observed when nucleic acids fold into a secondary structure. This phenomenon is due to the interaction of the π molecular orbitals of the nucleobases that occur when they are stacked one onto each other. For DNA duplex structures, hypochromicity is generally observed upon folding leading to an up to 30% decrease of the UV absorption at 260 nm, when compared to the unfolded form. For non-canonical structures, however, both hypo- and hyperchromicity can be observed. Observation of hypo- and hyperchromicity strongly depends on the considered wavelength.[160]

Measuring changes in absorption at a certain wavelength allows obtaining stability data by plotting it against temperature. The graphic representation of this variation is sigmoidal shaped, and it is called melting curve (*Figure* 2-8). The melting temperature T_m is a very common parameter used to compare the thermal stability of different nucleic acids and refers to the temperature at which half of the population of molecules is unfolded. This parameter can be obtained either by performing an annealing or a thermal denaturation of the oligonucleotide and it may vary depending on the oligonucleotide sequence, its concentration and the ionic strength of the media.

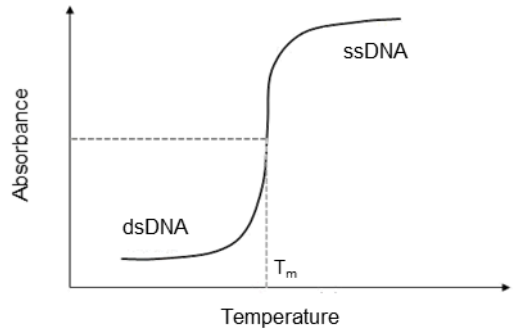


Figure 2-8. Graphic representation of a melting curve of an equilibrium between dsDNA and ssDNA.

Hysteresis provoked by different factors such as slow kinetics may lead to different T_m values for folding and unfolding processes. This phenomenon might be intrinsic to the sample under study, which commonly might indicate a complex equilibrium between multiple species or might as well occur if the experimental conditions are not properly adjusted. Smooth slope temperature ramps (typically 0.2 - 0.5ºC/min) minimize the risk of observing hysteresis in these experiments.

2.2.2. CIRCULAR DICHROISM (CD)[161,162]

Circular dichroism (CD) is a very helpful technique for the characterization of secondary structures of biomolecules. Linearly polarized light can be understood as a composition of two circular polarized constituents, one to the left and on to the right. When a beam of polarized light interacts with a chiral molecule, these two circular polarized constituents exhibit a differential absorption. This difference in absorbance of the left and right circular polarized light components provokes their resultant electric vector ($E_L + E_R$) to describe an ellipse. Ellipticity, which is the value of the tangent of the angle θ of the described ellipse, is commonly used to measure the circular dichroism of a sample. Experimentally, ellipticity is obtained by measuring the difference in absorption of the left and right circular polarized light components ($\Delta A = A_L - A_R$) and applying the following relation: $\theta = 32.98 \cdot \Delta A$. Molar ellipticity ($[\Theta_M]$, ($deg \cdot cm^2/dmol$)) is also used for comparison between samples at different concentrations and can be calculated as it follows, where θ_λ is the ellipticity expressed in mdeg and l is the optical path expressed in cm.

$$[\Theta_M] = \frac{\theta_\lambda}{10 \cdot l \cdot [oligo]}$$

For molecules to be sensitive to circular dichroism spectroscopy, they must absorb UV-visible light and have chirality. As explained in the previous section, nucleic acids fulfill the first requirement thanks to the presence of the nucleobases. As for the second, the residual chirality observed for unstructured nucleic acids is provided by nucleobase bonding to the chiral sugar units. However, the major contribution to the chirality of nucleic acids arises from the spectroscopic properties of their secondary structures. The CD signal is dependent of each wavelength so commonly CD spectra are performed in the wavelength range where the chromophores of the studied molecules exhibit absorption. In nucleic acids, CD is very sensitive to different nucleobase associations and intense CD bands appear depending on nucleobase interactions such as stacking. Thus, very characteristic spectra can be obtained to be used as fingerprints for the identification of secondary structures (see *Figure* 2-9). For i-motif structures, typical CD spectrum shows a maximum sharp band around 300 nm and a minimum broader band at around 240 nm.[162]

As mentioned, circular dichroism can also be used for obtaining thermodynamic data. Very similar to the melting experiments explained in the previous section, CD signal at a certain wavelength can be followed against a temperature gradient. In this case, CD signal always decreases as the structure is denatured. In this thesis, due to the pH-dependence of the i-motif structure, denaturation of the oligonucleotides was also performed by pH variation in order to obtain the pH value at which half of the molecules are unstructured (pH_T). In any case, CD denaturing experiments also lead to sigmoidal curves from which a T_m or pH_T values can be obtained as important stability data.

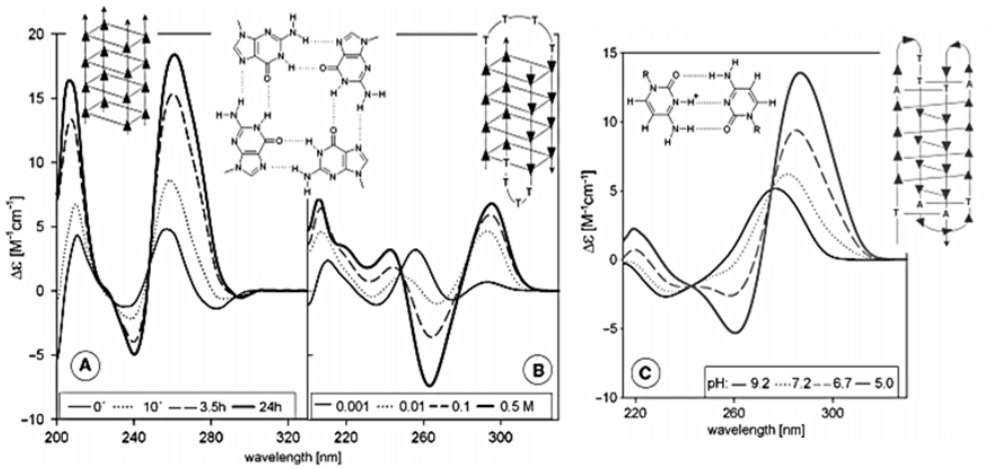


Figure 2-9. Characteristic CD spectra structural transitions for G-quadruplex (A and B) and i-motif (C) structures.[163]

2.2.3. FLUORESCENCE

Fluorescence is a luminescent property that exhibit certain molecules (fluorophores) or materials from which the absorption of electromagnetic radiation results in the emission of light through an excited transition state. Commonly, as one would expect, the energy of the emitted radiation is lower than the absorbed one, and this energy difference is measured by the quantum yield of the fluorescent substance. Quantum yield is defined as the ratio of the number of photons emitted to the number of photons absorbed thus giving the efficiency of the fluorescent process. Of special significance in the chemistry, biochemistry and biomedicine fields, as well as in nature, are the fluorescent substances that emit light in the visible region of the electromagnetic spectrum, allowing their visualization by the human eye.

In the context of structural studies of nucleic acids, the use of this technique becomes very useful as it can be strategically used for the detection and characterization of subtle structural transitions that other techniques are not able to trace. These strategies usually include the use of fluorescent analogues of nucleobases,[164] FRET systems[165] or other fluorescent probes.[166,167] Moreover, with the technological development of fluorescence microscopy, *in vivo* applications interestingly add value to the use of this technique[168,169] to the point of being used for the novel indirect visualization of DNA non-canonical structures *in vivo*.[46,170]

The physics behind fluorescence spectroscopy define the use of spectrofluorometers for the quantification of the fluorescent signal in *in vitro* experiments: the sample is irradiated using a high intensity light source that provokes the excitation of the fluorophore during a certain period of time. The incident light has typically been filtered through a monochromator in order to select the optimal

wavelength for the excitation. The fluorescent emission is again filtered through a monochromator before being recorded by a detector perpendicularly placed to the light source.

2.2.4. THERMODYNAMIC PARAMETERS

More than for simply obtaining T_m values from, melting curves can provide thermodynamic data for the folding/unfolding equilibrium process of a nucleic acid strucutre.[171] One of the approaches is the use of single melting experiments. In this case, two-state equilibrium between folded and unfolded species is required. Moreover, the molecularity of the reaction must be unequivocally determined.

The first step for obtaining thermodynamic parameters is to calculate the values of the equilibrium constant (*K*) at each temperature. *K* values arise from the following equation, which includes the molar fraction (x) of structured species at any given temperature, calculated from the variation of the absorption of the sample as a function of the temperature. Any other spectroscopic magnitude, as ellipticity or fluorescence can also be used to this purpose.

$$A = x \cdot A_{folded} + (1 - x) \cdot A_{unfolded} \quad \equiv \quad x = \frac{A - A_{unfolded}}{A_{folded} - A_{unfolded}} \qquad K = \frac{x}{(1 - x)}$$

Determination of thermodynamic data is ultimately achieved from a Van't Hoff plot, which fits the representation of ln*K* against 1/T to a linear model. Since *K* can be related to the Gibbs free energy as it follows, the application of this method can lead to the corresponding enthalpy and entropy values for the studied equilibrium. The accuracy of this approach greatly depends on whether ΔH^o and ΔS^o are temperature-independent, that is that $\Delta Cp^o = 0$ can be considered for the transition.

$$\Delta G° = \Delta H° - T \cdot \Delta S° = -RTlnK \qquad \equiv \qquad lnK = -\frac{1}{T} \cdot \frac{\Delta H°}{R} + \frac{\Delta S°}{R}$$

Usually, given that the representation of ln*K* against 1/T renders fairly linear plots. The chosen x values comprise a range of central values that afford more valid *K* values. These ranges comprise from 0.03<x<0.97 to 0.15<x<0.85 for more reliable results.

2.2.5. NUCLEAR MAGNETIC RESONANCE (NMR)

2.2.5.1. BACKGROUND AND FUNDAMENTALS

Together with X-ray crystallography, nuclear magnetic resonance (NMR) is the only technique capable of characterizing nucleic acids at atomic scale. Although crystallography was developed earlier and played a crucial role in the first structural discoveries of DNA, NMR offers nowadays several advantages that make of this technique an essential tool in the field. Due to technical limitations, NMR was originally

used to determine almost exclusively the structure of small organic molecules, but the development of superconducting high-field magnets, the introduction of multidimensional spectroscopy and the labeling strategies using isotopic atoms, have allowed NMR to be extensively used for the study of relatively big biomolecules such as oligonucleotides.[172] Unfortunately, for bigger nucleic acids (> 100 nu.), X-ray crystallography has still no competitor in terms of structural determination. One of the biggest differences between these two techniques is that NMR allows studying oligonucleotides in solution, a significantly more similar media than that of the physiological one. The other significant advantage is that NMR allows performing dynamic studies that are of high importance to explore, *e.g.*, flexible regions of nucleic acids involved in recognition processes. Remarkably, related to the visualization of i-motif structures in cellular medium, it has been reported very recently that NMR spectroscopy can be effectively used to assess the formation of the motif, and to provide information regarding its stability under physiological conditions.[63]

NMR spectroscopy fundamentals are based on the fact that randomly oriented spins from magnetically active nuclei can be oriented under a static magnetic field. This orientation can occur, for nuclei with spin values of ½ (typically ^{1}H, ^{13}C, ^{15}N, ^{19}F or ^{31}P), in the same or in the opposite direction of the magnetic field and thus, two populations (α and β) of differently oriented nuclei with different energies can be distinguished. Generally, the α population exhibits lower energy than the β in equilibrium. The detection of the NMR signal is achieved after the perturbation of the oriented spins by an oscillating magnetic field, also known as resonance pulse, which provokes the precession of the magnetized spins. Due to the small energetic difference between α and β states, NMR spectroscopy is less sensible than other spectroscopic techniques and high concentrated samples (mM) are required. Nonetheless, since the energetic gap between α and β populations depends on the intensity of the magnetic field, stronger magnets yield better resolved spectra.

Simply from the previous information, one could question the effectiveness of NMR spectroscopy assuming that all the nuclei of the same nature would equally resonate depending on the applied magnetic field. However, what makes NMR spectroscopy so interesting is that the resonating frequency of a certain nucleus is sensitive to the magnetic field provoked by surrounding electrons. This phenomenon is known as shielding and its influence is reflected in the chemical shift of the signals which allows the use of NMR spectroscopy for structural elucidation of molecules. Essentially, highly shielded nuclei receive an effective lower magnetic field which translates into a lower energetic gap between the two populations. This effect is commonly referred as upshifting. On the other hand, when talking about a downshifted nucleus, we refer to remarkably unshielded atoms that receive a higher effective field. For i-motif structures, this explains why the characteristic signal of the proton corresponding to the hemiprotonated $C{\cdot}C^{+}$ base pair shows such a higher chemical shift (15 – 16 ppm).

Another important feature of NMR spectroscopy that is crucial for its applications in the structural chemistry field is the splitting patterns that the signals exhibit due to the effect that provokes on a

resonating nucleus the resonance of neighbor nuclei scalarly bound to it. This phenomenon is called coupling effect and, in the field of structural determination of nucleic acids, is useful for the determination, for instance, of the conformation of sugar rings.

2.2.5.2. NMR SPECTROSCOPY FOR THE CHARACTERIZATION OF NUCLEIC ACIDS[173]

2.2.5.2.1. 1D spectra

Generally, the first step for the structural elucidation of a macromolecule involves the acquisition of one-dimensional ^{1}H NMR spectra. Although the complete assignment of the signals is nearly impossible due to the complexity and overlapping of the signals, 1D NMR spectra are useful to confirm certain structural features, generally from the exchangeable proton region of the spectrum. Moreover, information on the formation of alternative structures or on the symmetry or molecularity of the structure can also be obtained from one-dimensional experiments. In terms of stability, 1D NMR also results very useful for the qualitatively determination of the stability of the structure by performing different experiments under different conditions such as temperature, concentration or pH. Especially for i-motif structures, its pH-dependent nature, allows to follow the folding/unfolding of the motif by specific 1D NMR signals.

2.2.5.2.2. 2D spectra

To solve the structure of macromolecules such as nucleic acids or proteins, more complex experiments than simply 1D NMR are needed. Two-dimensional NMR spectroscopy, also called correlation experiments, generally overcome overlapping problems and allow to establish nuclei connections that are key for obtaining structural information. The general pulse sequence for 2D NMR experiments is exemplified in *Figure* 2-10 and consists of different steps:

1. Preparation: initial magnetization of the nuclei through a set of radiofrequency pulses.
2. Evolution: short period of time in which the nuclei undergo precession as no pulses are delivered.
3. Mixing: alteration of the magnetization by delivery of another set of pulses.
4. Detection: observation of the relaxation of the nuclei as a function of time.

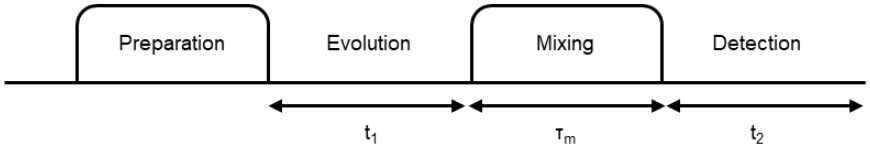


Figure 2-10. Schematic representation of the general pulse sequence in two-dimensional NMR spectroscopy.

Different 2D experiments can be performed depending on the type of pulses delivered to provoke magnetic alteration during the mixing time. These different experiments allow correlating atoms that fulfill specific conditions (scalarly bound, close in the tridimensional structure, etc.) which results very useful for obtaining structural information. The following homonuclear ^{1}H-^{1}H two-dimensional experiments are the most used in this thesis for the characterization of i-motif structures:

Total Correlation Spectroscopy (TOCSY).[174] This experiment uses potent electromagnetic pulses to block the magnetization during the mixing time in order to fulfill the Hartmann-Hahn condition (magnetization transfer between two scalarly coupled spins occurs when they have the same effective precession frequency). TOCSY experiments are useful to correlate all the protons belonging to the same spin system since the magnetization transfer occurs between atoms scalarly coupled.

Nuclear Overhauser Effect Spectrocopy (NOESY).[175] Based on the nuclear Overhauser effect, NOESY experiments correlate the magnetic diploes of proximal atoms. After the first magnetization of the system, a second pulse alters one of the magnetic components from the equilibrium. During the mixing time, no pulses are applied, and the system tends to recover the equilibrium situation via dipolar relaxation. A final electromagnetic pulse transforms the transferred magnetization in order to be detected.

Very interestingly for the structural elucidation of macromolecules, the intensity of the NOE peaks can be directly related to the distance between the correlated atoms. A relatively rough approximation relates, for small mixing times, the intensity of the NOE peaks with the inverse of the sixth power of the distance between the two correlated atoms. The intensity of the NOE peaks is dependent, among other factors, on the mixing time during the pulse sequence. The sensitivity of the experiment usually allows correlating atoms within a range up to approximately 6Å.

2.2.5.2.3. Assignment[176,177]

The general assignment strategy combines scalar correlation experiments (TOCSY) to identify different spin systems, and dipolar correlation experiments (NOESY) to assign specifically residues from the nucleic acid chain. The methodology is based on strategic contacts observed in B DNA, so it must be

used carefully when studying non-canonical structures since chemical shift values might vary due to a different tridimensional arrangement. The different resonances corresponding to the different kinds of non-exchangeable protons are quite well separated in terms of chemical shift except for the H4'/H5'/H5" protons which are generally difficult to assign. The following table shows the typical proton chemical shifts for the different nucleotides, including those involved in non-canonical base pairing that are relevant for this thesis:

ATOM	THYMINE	CYTOSINE	ADENINE	GUANINE
H1'	5.0 – 6.0	5.0 – 6.0	5.0 – 6.0	5.0 – 6.0
H2'	1.7 – 2.3	1.7 – 2.3	2.3 – 2.9	2.3 – 2.9
H2"	2.1 – 2.7	2.1 – 2.7	2.4 – 3.1	2.4 – 3.1
H3'	4.4 – 5.0	4.4 – 5.0	4.4 – 5.2	4.4 – 5.2
H4'/H5'/H5"	3.8 – 4.3	3.8 – 4.3	3.8 – 4.3	3.8 – 4.3
H6/H8	6.9 – 7.9	6.9 – 7.9	7.7 – 8.5	7.7 – 8.5
CH_3/H5/H2	1.0 – 1.9	5.0 – 6.0	7.7 – 8.5	-
NH (WC)	13.0 – 14.5	-	-	11.5 – 13.5
NH (unpaired)	10.0 – 11.0	-	-	10.0 – 11.0
NH (GT)	10.0 – 12.0	-	-	10.0 – 12.0
NH (C^+)	-	15.0 – 16.0	-	-
NH_2 (WC)	-	6.5 – 8.5	7.0 – 8.0	8.0 – 9.0
NH_2 (unpaired)	-	6.7 – 7.0	5.0 – 6.0	5.0 – 6.0
NH_2 (C^+)	-	8.0 – 9.5	-	-

Table 2-1. Typical chemical shift values of ^{1}H NMR in DNA (ppm).

A common starting point is to identify each spin system and to determine to which nucleobase belongs to. This can be achieved by identifying the scalar correlation cross-peaks from TOCSY experiments corresponding to the H5 and H6 protons of the cytidine or to the methyl group and the H6 protons of the thymidine. The second step is to relate each base with its own sugar and to determine its position in the oligonucleotide chain. This can be done by following the NOE cross-peaks between the base and its own sugar and the sugar of the previous nucleotide in the 5' – 3' sense as it is shown in *Figure* 2-11. What usually results more accessible in order to determine the sequence of the oligonucleotide, is to follow the H6/H8 interactions with the H1'/H2'/H2" protons of the sugar. The stereospecific assignment of the H2'/H2" protons can be done by comparing the intensity of the NOE cross-peaks between these protons and the H1': The cross-peak between H1' and H2" (cis) is more intense than that found for the interaction between H1' and H2' (trans). While the H3' protons are commonly easy to differentiate due to their chemical shift and the intensity of their cross-peaks with the H1'/H2'/H2" residues, the H4'/H5'/H5" atoms are harder to specifically assign due to overlapping of their peaks. For the assignment of these protons, ^{31}P-^{1}H correlation experiments are extremely useful.

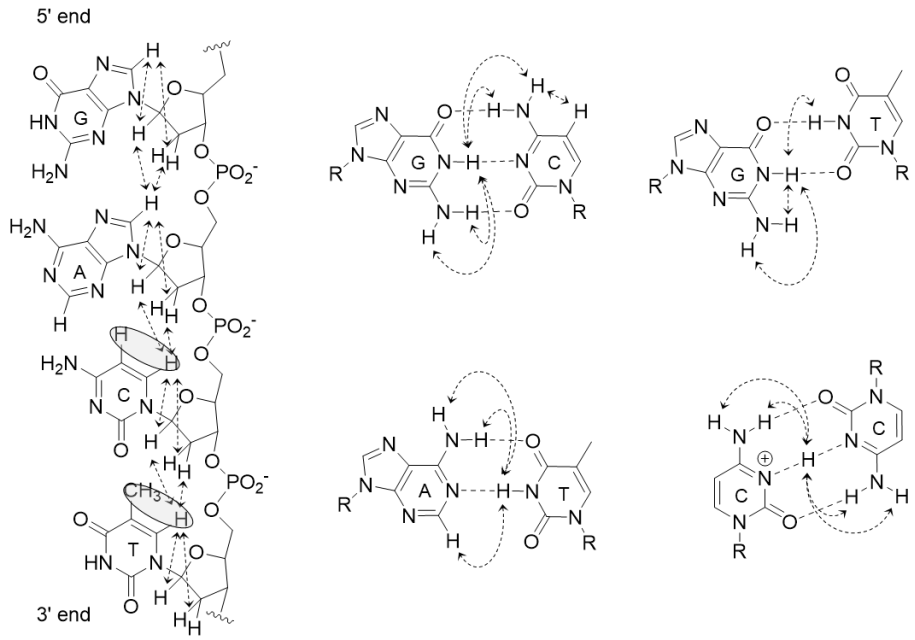


Figure 2-11. Main contacts used for 2D ^{1}H NMR assignment of oligonucleotides.

The NOESY spectra in H_2O:D_2O 9:1 allow the assignment of the exchangeable protons by following the amino – imino and imino – imino cross-peaks between strands. For WC base pairs (*Figure* 2-11), the G·C base pair can be easily assigned following the cytidine's H5 cross-peak with its own H4 amino protons and the N1 imino proton of the guanidine. For A·T base pairs, the cross-peak between adenine's H2 proton and the N3 imino proton of the thymidine can be also clearly identified. The i-motif characteristic exchangeable protons cross-peaks are those between the protonated imino N3 proton and the H4 amino protons of the cytidine (*Figure* 2-11).

Besides from the previously mentioned NOE signals of the heavily downshifted imino proton of hemiprotonated C:C$^+$ base pairs, i-motif structures exhibit other characteristic NOE contacts that are crucial for the determination of the topology of the motif. These contacts are represented in *Figure* 2-12. For instance, as it was mentioned in the introduction chapter, inter-strand H1'-H1' contacts (purple in *Figure* 2-12) through the narrow grooves of the structure are observable due to proximity of the sugar-phosphate backbone. On the other hand, contacts through the major groove occur majorly between nucleobase protons. However, most interestingly, H41/H42-H2'/H2" contacts (red in *Figure* 2-12) are very useful for the determination of the stacking order of the C:C$^+$ base pairs since they only occur in back-to-back steps between the 3' sides of the residues.[56]

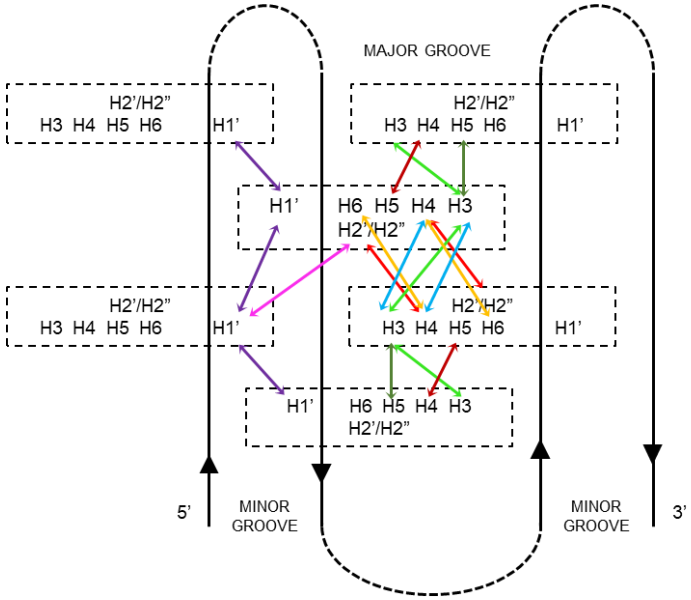


Figure 2-12. Typical non-sequential NOE contacts potentially observed in i-motif structures.

2.2.5.2.4. Calculation of solution structures

The correlation between atomic coordinates and experimental restrictions allows obtaining three-dimensional solution structures of biomolecules. Generally, from this process several structures that match the experimental information are obtained and thus, an average structure is represented from the sum of valid conformations obtained as a result from the structural determination based on NMR experiments.

2.2.5.2.4.A. Experimental restrictions

The most important experimental data used for the calculation of solution structures is obtained from the intensity of the NOE peaks, which not only depends on the proximity of the nuclei, but also to the rigidity of the molecule, the mixing times utilized in the NMR experiment and the neighboring atoms. The general approximation applied to the determination of spatial restraints is to consider the molecule as a rigid solid and that each pair of protons is isolated from the rest. Taking into account these considerations, the intensity of the NOE peaks is proportional to the mixing time of the experiment and inversely proportional to the sixth potency of the distance between the protons.

$$NOE \propto \frac{t_m}{r^6}$$

The proportionality constant is calculated from NOE cross-peaks between fixed atoms such as H2'-H2" or aromatic protons. The drawbacks of this simplification are that is only valid for short mixing times, that allow ignoring the spin diffusion effect, and that the consideration of a biomolecule as a rigid solid is generally unrealistic since most structures exhibit very active and complex solution dynamics.

In practice, a preliminary qualitative classification is performed based on a rough approximation on the intensity of the NOE cross-peaks: high, medium and low (each assigned to a range of distances). In globular-shaped structures such as proteins, this rough approximation is compensated with the high number of cross-peaks between protons at long distances within the sequence that arise from the intrinsic topology of the structures. On the other hand, in other biomolecules like nucleic acids, these contacts are less common, and the estimation of experimental distances must be refined.

Such improvement can be achieved by applying the complete relaxation matrix concept.[178,179] This theoretical model takes into consideration all the relaxation processes of all the protons within a structure, simulating the spin diffusion effect. This model has an experimental limitation since typically not all the protons can be unequivocally assigned and thus some spots remain unfilled in the construction of the matrices. This limitation is fixed by calculating a preliminary structure to fill the experimental gaps and from this point an iterative calculation affords the final constraints (within a range) for the calculation of the three-dimensional structure. These iterations are performed using the MARDIGRAS software.[180] As a special consideration, NOE cross-peaks from exchangeable protons are less intense than they should due to the effect of the solvent.

2.2.5.2.4.B. Restrained molecular dynamics (RMD)

Based on the mathematical approximation they use, different calculation methodologies can be defined. The best-known and most utilized among these methodologies is the use of restrained molecular dynamics. However, another methodology based on the minimization of a target function can be very useful for obtaining preliminary structures This methodology is focused on the definition of a positive function that is equal to zero when the distance restrictions are accomplished and different to zero if not. Therefore, obtaining of a solution structure that matches the experimental restrictions is reduced to the minimization of such function. The steepest descent methodology operates in the space of the torsion angles of the biomolecule and searches iteratively the minimum value for the function in the same direction of the gradient. For the structures calculated in this thesis, the DYANA software has been used for the calculation of preliminary structures based on molecular dynamics algorithms.

From these preliminary structures, restricted molecular dynamics (RMD) can be applied. This methodology uses Newton's equations for a system of particles. These equations include the preliminary coordinates previously obtained and an empirical potential factor that includes bonding interactions, non-bonding interactions and the experimental restrictions as an exponential factor in order to subjugate

the optimization to the experimental data. The complications behind this methodology arise from the excess or the lack of importance given to this exponential factor, which would lead to the violation of theoretical constraints or to the negligence of experimental data, respectively. RMD protocols follow a simulated annealing procedure in order to overcome energetic barriers and local minimum wells.

2.2.5.2.4.C. Analysis of the structures[181]

Obtaining well-defined three-dimensional structures is tightly related to the amount, the quality and the distribution of the experimental restrictions. Ideally, a notable definition of a structure includes non-less than fifteen restrictions per nucleotide.

The analysis of the quality of the structures obtained from RMD methods includes:

- Determine the fidelity between the obtained structures and the experimental data obtained from NMR experiments. A direct indicator of this fidelity is the value of the final experimental potential energy: the lower the better. However, this potential includes several variables that might mislead to incomplete conclusions. Therefore, another analysis based on the sum of residual violations is commonly performed. Nevertheless, this is an indirect analysis based on the distances between the atoms and not on the correlation with the intensity of the NOE cross-peaks.

- The analysis of the degree of definition of the resultant structures. This is achieved from the calculation of the root square mean deviation (RSMD) of the group of obtained structures, which also acts as a good indicator of the accuracy of the results:

$$RSMD = \sqrt{\frac{\sum_{i=1}^{n} |v_i - w_i|^2}{n}}$$

- Finally, the analysis of the structural results like dihedral angles and helicoidal parameters in order to compare them to canonical well-defined structures. Remarkably inconsistent experimental results might reveal potential inconsistencies of the calculation. The software CURVES allow the calculation of these structural parameters.[182]

3. INCORPORATION OF psC IN i-MOTIFS

3.1. BACKGROUND AND OBJECTIVES

Understanding the factors that potentially enable the formation of i-motif structures under physiological conditions is imperative in order to control their dynamic pH-dependent behavior. A considerable amount of chemical modifications differently used in i-motif structures were described in the introduction of this thesis, showing how some strategical approaches led, not only to significantly enhanced stabilities of the motifs, but more importantly to the comprehension of which molecular forces and interactions are crucial for the formation and stabilization of the i-motif.[97,100,106]

The present chapter is focused on exploring the use of pseudoisocytidine (psC), a neutral analogue of protonated cytidine, to reproduce the formation of C:C$^+$ base pairs in a pH-independent manner. The early steps of this study were performed in previous master's degree projects performed by X. Solés and me.[183,184] Thus, this introductory section will englobe the preliminary results that were obtained. In the following sections, a deeper structural discussion based on 2D NMR experiments and restrained molecular dynamics calculations will be presented, as well as an extended study in which the human telomeric sequence was included.

3.1.1. NEUTRAL ANALOGUES OF PROTONATED CYTIDINE

Pseudoisocytidine is a cytidine isostere whose connection between the nucleobase and the ribose ring occurs through a C-C bond instead of the natural C-N glycosidic bond. Also, psC exhibits a tautomeric equilibrium that involves a proton exchange between N1 and N3 nitrogen atoms that must be taken under consideration since it can be relevant for predicting its hydrogen bond pattern with other nucleobases. Interestingly, pK_a values for psC (pK_a (N1) = 3.79 and pK_a (N3) = 3.69) are lower than that of cytidine (pK_a (N3)= 4.5), indicating that psC protonation requires lower pH conditions.[185]

The use of cytidine analogues for reproducing the neutral formation of protonated nucleobase associations has been previously tested in parallel triplexes. Different studies utilizing 6-oxocytidine[186] or pseudoisocytidine,[187–189] proposed the construction of stable triplex structures at neutral pH with strategic replacements of cytidine residues involved in the formation of C$^+$-G:C triplets. Since the phosphoramidite derivative of psC is commercially available, its introduction in i-motif structures is a straightforward process by standard oligonucleotide synthesis methodology (see *Section* 8.5.1 in Methods).

The 3*H*-tautomer of psC, which has been theoretically found as the most stable,[185] could form a neutral psC:C base pair with cytidine. Thanks to the extra hydrogen-bond donor in the N3 position of psC, psC:C base pair would be stabilized by three hydrogen bonds between the nucleobases, yielding a base pair completely isomorphic to the unmodified C:C$^+$. The formation of psC:psC base-pairs from the association of the two different psC tautomeric forms should also be considered (*Figure* 3-1). Thus,

introduction of psC in i-motif structures is expected to increase the number of hydrogen-bonds in i-motif structures without the energetic cost required for cystosine protonation. This effect might provoke an overall stabilization of i-motif structures at neutral pH. However, the lack of the positive charge inherent to C:C+ base pairs might have a detrimental effect on the stability. Interestingly, an analogous study also aiming to the formation of unprotonated i-motif structures performed by Wright *et al*., tested the incorporation of 2'-deoxyriboguanylurea in the human telomeric i-motif.[190] They report that although 2'-deoxyriboguanylurea can form a theoretical neutral base pair with cytidine, the resulting i-motif structures are less stable than those formed by the natural unmodified sequence.

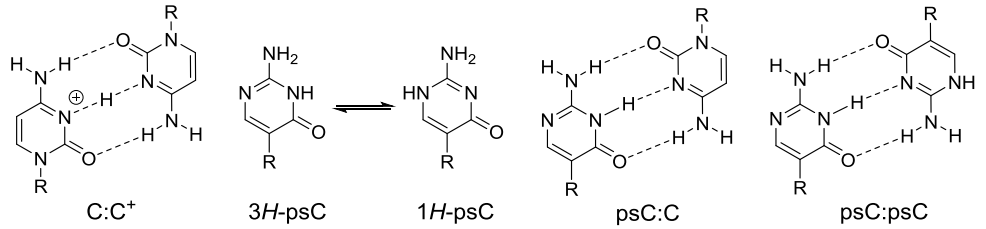


Figure 3-1. Hemiprotonated C:C$^+$ base pair, pseudoisocytidine tautomeric equilibrium and possible nucleotide associations involving either cytidine or pseudisocytidine residues expected in the structures studied in this chapter.

3.1.2. SEQUENCES

In order to explore the effect of psC on i-motif formation, a dimeric i-motif structure that was previously characterized in detail by our research group was chosen. This structure, named mini i-motif, as it only exhibits two C:C$^+$ base pairs in its core, was described from the 9-mer non-cytosine rich sequence d(TCGTTTCGT), named **M**, and its cyclic 10-mer analogue.[191] The unexpected high stability of this dimeric i-motif is most likely caused by the stacking interactions provided by the formation of G:T:G:T minor groove tetrads that flank the ends of the cytosine core. As it was described in the introductory chapter of this thesis, the formation of these type of tetrads occurs from the association of two G:T mismatches through their minor groove side, enabling extra hydrogen-bonds formation between guanine residues (see *Figure* 3-2).[191,192] Because of its dimeric nature, two different topologies for the **M$_2$** dimer can be observed in equilibrium from the different association of **M** moieties: the head-to-tail and the head-to-head dimers (*Figure* 3-2).

Due to its unique structural features, the mini i-motif structure has some limitations for exploring the full scope of psC incorporation. Although C2 and C7 are not equivalent residues, given the highly symmetrical nature of the three-dimensional structure of this i-motif, these residues have near identical chemical environment in the structure and no large differences in their chemical shifts should be expected between replacing C2 or C7 by psC. Consequently, an 11-mer sequence derived from **M**, was designed with the aim of obtaining a dimeric i-motif structure with an extended core of four C:C$^+$ base

pairs and also capped by G:T:G:T minor groove tetrads, (**MC2**, d(TCCGTTTCCGT)). This construction may allow us to evaluate different chemical environments for the incorporation of psC. The resultant **MC2$_2$** dimer was fully characterized, affording stable i-motif structures with the previously described features and both head-to-tail and head-to-head topologies in slow equilibrium (*Figure* 3-2).[183] In **MC2$_2$** structures, hemiprotonated cytosine residues in different chemical environments are found, allowing a complete study of the effect of psC residues in different positions of i-motif structures.

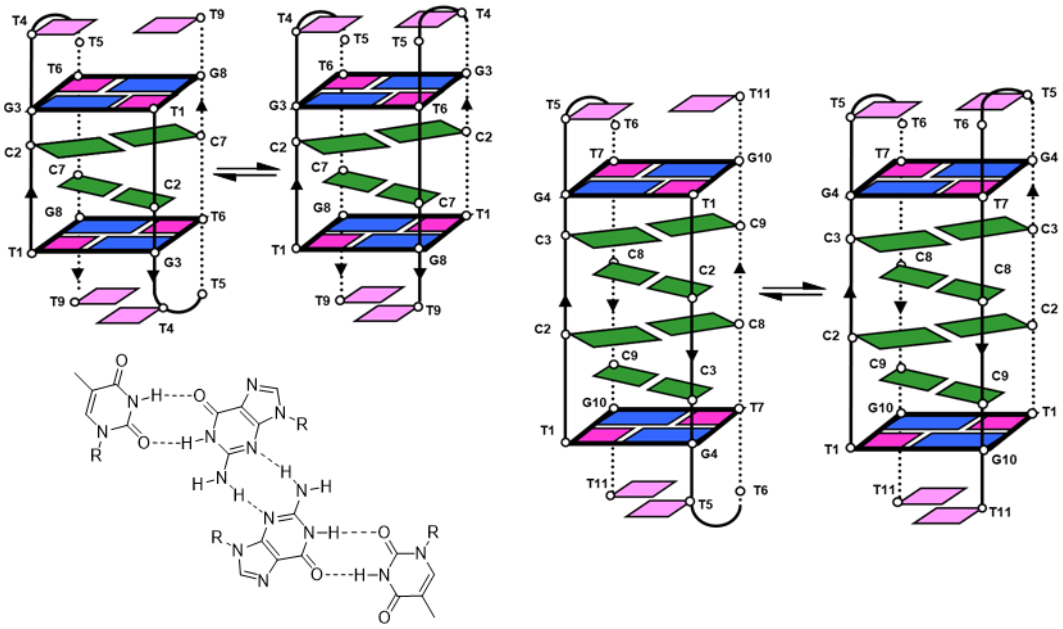


Figure 3-2. Schematic representation of the head-to-tail and head-to-head topologies of the dimeric structures of **M$_2$** (top left) and **MC2$_2$** (right). Minor groove G:T:G:T tetrad (bottom left).

Therefore, psC was incorporated replacing cytidine residues in position 7 in **M** (**M_psC(7)**) and in positions 8 or 9 in **MC2** (**MC2_psC(8)** and **MC2_psC(9)**, respectively) see *Table* 3-1). The presence of psC in different positions of **MC2**, offers the possibility of analyzing the effect of cytidine replacement in diverse structural environments. Depending on the stacking order of the C:C$^+$ tract, psC residues will be found immediately stacked below/above the G:T:G:T tetrads, or placed in the center of the C:C$^+$ stack. Moreover, for both **M** and **MC2** psC-containing sequences, we expected to observe whether a particular topological configuration is preferred against the other. This preference was assumed to be provided by the relative stability of the different possible base pairs (*i.e.*, C:C$^+$, psC:C or psC:psC). While head-to-head orientations would imply the formation of C:C$^+$ and psC:psC homopairs, psC:C heteropairs would be found only in head-to-tail alignments. These three psC-containing sequences were prepared during the master's degree projects mentioned above.[183,184]

3.2. STABILITY RESULTS

The preliminary experiments performed during both master's degrees afforded important stability data for the previously mentioned sequences. 1D ^{1}H-NMR and CD techniques were utilized to assess i-motif formation and to evaluate the stability of each sequence against thermal and pH denaturation.

As an outline of the conclusions obtained from these experiments, it was revealed that all three psC-containing sequences were able to fold into i-motif structures. However, as observed from 1D ^{1}H-NMR studies shown in *Figure* 3-3, the only psC-containing structure that remained stable under neutral pH conditions was the one formed by the sequence **MC2_psC(9)**. Multiple imino signals around 15 ppm, characteristic of i-motifs, were also found for **M_psC(7)** and **MC2_psC(8)** sequences, but only under acidic conditions.

When compared with the non-modified sequences, the incorporation of a psC residue replacing C7 in **M** clearly destabilizes mini i-motif formation, whereas cytidine substitutions in **MC2** seem to have site-specific effects. According to NMR data, formation of **MC2** (and also **MC2_psC(8)**) i-motif structures requires acidic conditions, whereas for **MC2_psC(9)** protonated imino signals are observable at pH 7, indicating an enhanced i-motif stabilization at neutral pH for this sequence. In the case of unmodified sequences, **M** or **MC2**, head-to-head or head-to-tail species are similarly populated at all pH values.[76,183,184] In contrast, in psC-containing sequences the number of signals and their relative intensities change upon lowering the pH, indicating that the relative stability of both orientations is not the same at different pH values.

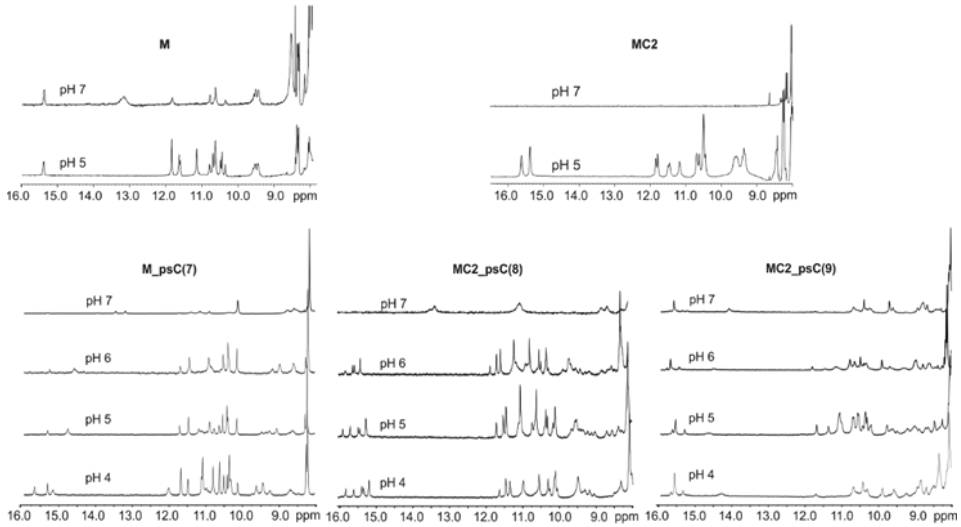


Figure 3-3. 1D ^{1}H-NMR spectra at different pH. H_2O/D_2O 90:10, 25 mM phosphate buffer, 100 mM NaCl, T = 5 °C, [oligonucleotide] = 0.7–1 mM.

The values of transitional pH midpoint were calculated from pH titration curves followed by CD spectroscopy. Experimental T_m values were analogously obtained from thermal denaturing CD experiments (see *Section* 8.7.3 in Methods). Results for the 9-mer sequences showed that **M_psC(7)** exhibits a lower pH_T value than its unmodified analogue. Furthermore, at low oligonucleotide concentration, the structures from both sequences exhibit a very low (or marginal) thermal stability at neutral and even mildly acidic pH. In the case of the 11-mer sequences, the overall stability of the i-motif structures seems to be enhanced by the extended $C:C^+$ tract under these conditions. More in detail, the psC-containing sequences showed different stability compared to **MC2**. Remarkably, among all the sequences studied in this section, **MC2_psC(9)** exhibits the higher pH_T value (6.7) and is the only structure that remains structured at pH 7, although presenting a low T_m value (16.7 ºC). On the other hand, substitution in position 8 renders a less stable structure than that of the unmodified oligonucleotide. This data reinforces first evidence of the different effect that produces psC on the i-motif structure depending on which position occupies, observed in the 1D 1H-NMR spectra (*Figure* 3-3).

NAME	SEQUENCE	pH_T	T_m @pH5.5	T_m @pH7
M	d(TCGTTTCGT)	5.8 ± 0.1	~ 10	-
M_psC(7)	d(TCGTTT***psC***GT)	5.3 ± 0.1	-	-
MC2	d(TCCGTTTCCGT)	6.5 ± 0.1	29.6 ± 0.5	-
MC2_psC(8)	d(TCCGTTT***psC***CGT)	6.1 ± 0.1	28.5 ± 0.5	-
MC2_psC(9)	d(TCCGTTTC***psC***GT)	6.7 ± 0.1	37.4 ± 0.5	16.7 ± 0.5

Table 3-1. Sequences and stability data against pH and temperature (ºC) obtained from CD experiments. Experimental conditions: [oligonucleotide] = 20 µM, 25 mM of buffer solution (cacodylate (pH 5.5) or phosphate (pH 7)) and 100 mM NaCl.

3.3. STRUCTURAL CHARACTERIZATION

The complete characterization of the structures was achieved from 2D NMR experiments. The full assignment of the NOESY spectra was complicated by the presence of mixtures of i-motif species for most sequences.

NMR assignments were conducted in all cases on the basis of the assignment of the unmodified sequence, **M**, previously reported.[76,183,184] Some key features are common to all the spectra, as the characteristic imino signals of hemiprotonated $C:C^+$ base pairs observed at 15-16 ppm, or the number of their cross-peaks with cytosine amino signals (H41/H42). Also, the number of aromatic spin systems in the spectra is very informative, since provides key structural information on the number of coexisting species. It must be noted that in all the cases studied in this chapter, the equilibria between different species are slow in the NMR time scale. Thus, different signals can be observed for the same proton in each of the exchanging species. Although, severe overlapping is found in the non-exchangeable spectra, in general, exchangeable proton regions exhibit very little overlapping and the cross-peaks

patterns of imino and amino signals provide a wealth of structural information, allowing the assignment of many of these signals to one of the co-existing species.

Finally, it must be noted that the i-motif structures studied in these sections are symmetric dimers, and the nuclei belonging to symmetry related residues are equivalent from the NMR point of view.

Due to its higher stability and its formation at pH 7 assessed by 1D NMR, we focused on the structure of **MC2_psC(9)** for its characterization. Nonetheless, an outline of the different structures and topologies adopted by **MC2_psC(8)** and **M_psC(7)** was proposed.

3.3.1. MC2_psC(9)

3.3.1.1. **MC2_psC(9)** FOLDS INTO A DIMERIC HEAD-TO-TAIL i-MOTIF AT pH 7

Among the proposed psC-containing sequences, **MC2_psC(9)** is the only one that exhibits i-motif structuration at neutral pH. NOESY spectra of **MC2_psC(9)** at pH 7 shows a set of signals and spin systems consistent with a unique major species and topology. Exchangeable protons region corresponding to imino and amino protons are shown in *Figure* 3-4.

The proposed topology corresponds to a head-to-tail dimer, since it is the only association that incorporates the formation of psC:C base pairs in the i-motif structure. The identification of the psC residue spin system is easily achieved by following the imino-amino NOE cross-peaks that exhibit no further connections with H5 protons. Then, the psC9:C3 hetero base-pair formation is confirmed, as well as the C2:C8 hemiprotonated base pair, through imino-amino cross-peaks between non-equivalent residues as a unique imino signal shows cross-peaks with two pairs of amino protons (*Figure* 3-4). Remarkably, the imino proton of psC in psC:C is shifted upfield (14.05 ppm) compared to the imino proton of the $C:C^+$ base pair (15.53 ppm).

Analogously to its unmodified analogue, **MC2_psC(9)** also prefers the stacking order that allows the stacking of guanidine residues with their sequentially immediate cytidine (or psC) residues. The paths for connecting both C3-G4-T5 and psC9-G10-T11 fragments can be assessed from the following NOE cross-peaks: H41/H42C3/psC9-H1G4/G10 and H1G4/G10-MeT5/T11. The stacking order of the core of the structure is directly determined from amino-H2'/H2" back-to-back (3'-3') contacts between stacked residues through the major groove. These contacts are observed in the NOESY spectra and were assigned to occur between equivalent residues (*i.e.* H42-H2'/H2"C8), see *Figure* 3-5. Finally, the connection between the $C2:C8^+$ core and the C3:psC9 outer pairs is confirmed via $H3^+$C2:C8-H3psC9 imino-imino cross-peaks. Thus, the 3'-E topology of the **MC2_psC(9)$_2$** head-to-tail dimer was established by following the path T11-G10-psC9-C2-C8-C3-G4-T5.

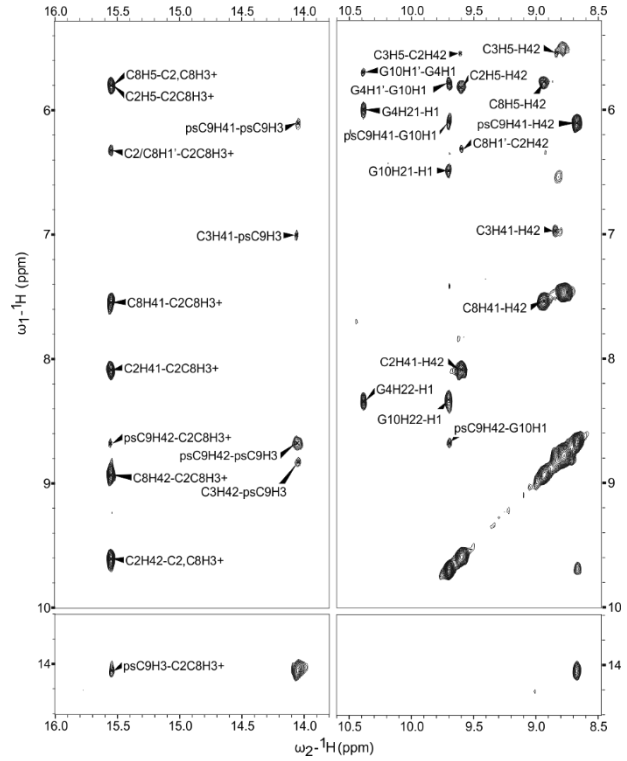


Figure 3-4. Imino and amino exchangeable protons region of NOESY spectrum (250 ms) of **MC2_psC(9)**. H_2O/D_2O 90:10, 25 mM phosphate buffer pH 7, 100 mM NaCl, T = 5 °C, [oligonucleotide] = 0.8 mM.

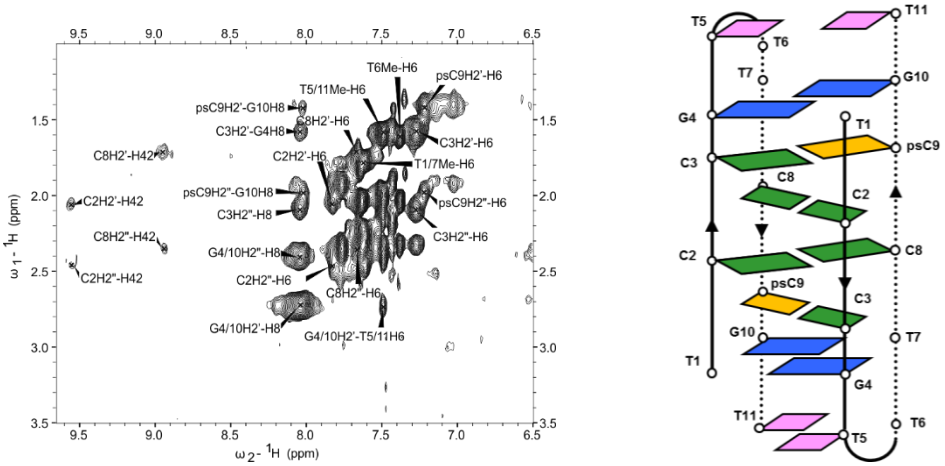


Figure 3-5. Amino/Aromatic-H2'/H2" protons region of NOESY spectrum (250 ms) of **MC2_psC(9)**. H_2O/D_2O 90:10, 25 mM phosphate buffer pH 7, 100 mM NaCl, T = 5 °C, [oligonucleotide] = 0.8 mM (left). Schematic representation of the head-to-tail major structure of **MC2_psC(9)** at pH 7 (right).

Unexpectedly, the formation of G:T:G:T minor groove tetrads could not be assessed. However, whereas H1G-H3T imino-imino NOE signals were not observed to confirm G:T base pair formation, H1'G4-H1G10 and H1'G10-H1G4 cross-peaks reveal the association of guanine residues through their minor groove side to form G:G base pairs. The schematic representation of the described structural features of **MC2_psC(9)** at neutral pH is shown in *Figure* 3-5.

3.3.1.2. SOLUTION STRUCTURE OF **MC2_psC(9)** AT pH 7

The solution structure of **MC2_psC(9)** at neutral pH was obtained based on a total of 128 distance restrictions. Some of the more relevant contacts are depicted in *Figure* 3-6).

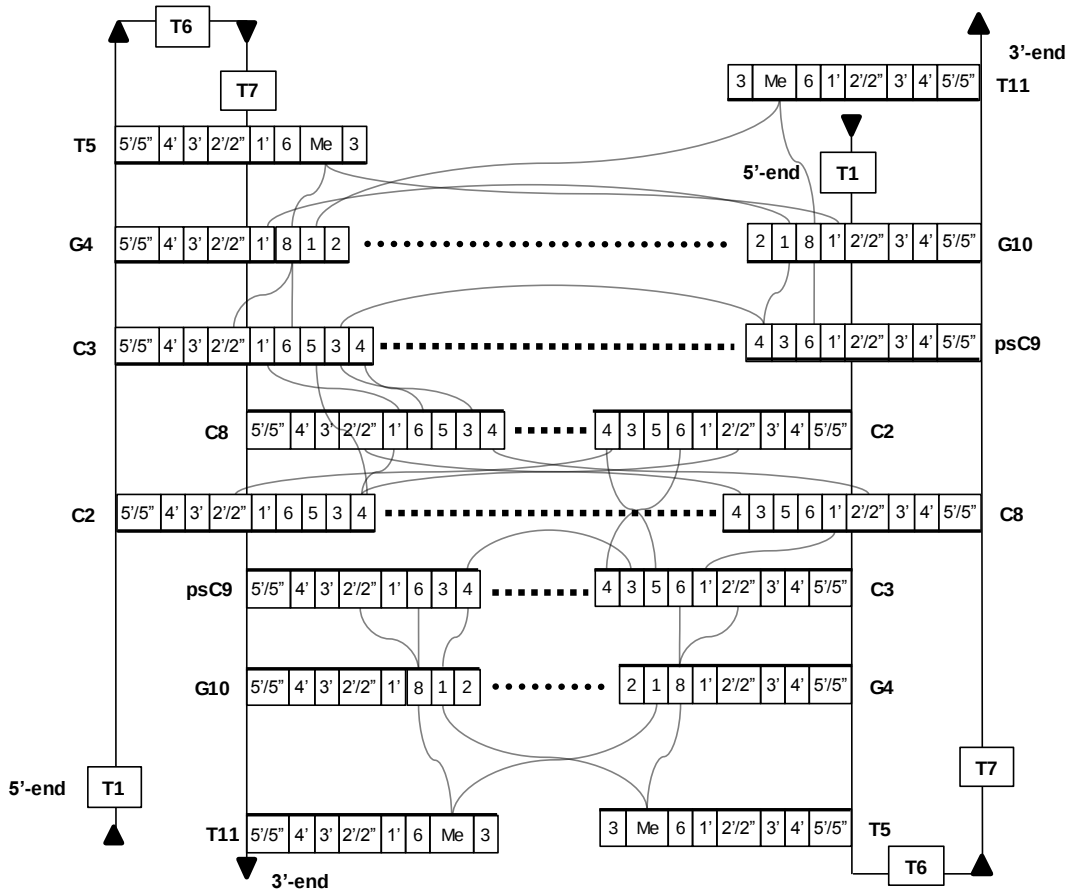


Figure 3-6. Scheme of the most relevant NOE contacts. Round dotted lines indicate G:G base pairs and square dotted lines stand for C:C$^+$ base pairs. 5'-sides of the nucleobases have been represented with thicker lines.

An ensemble of 10 structures was obtained through restrained molecular dynamics calculations (see *Section* 8.10 in Methods) (*Figure* 3-7, PDB code 5NIP). Except for the thymine loop residues, the structures are well-defined with a root-mean-square deviation (RMSD) value of 0.8 Å. *Table* 3-2 shows detailed statistic data regarding the calculations of the structure.

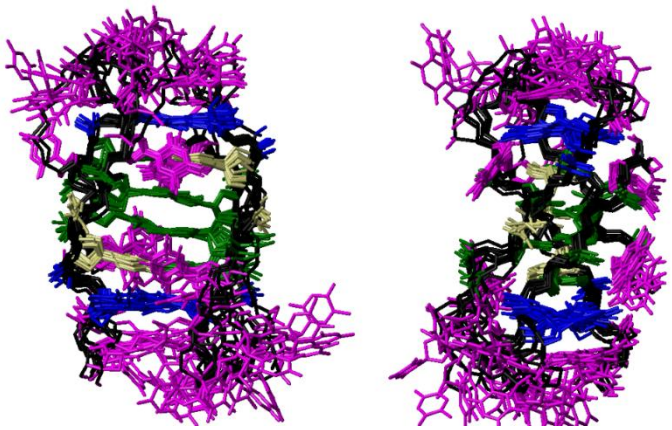


Figure 3-7. Two views of ensembles of ten structures of **MC2_psC(9)** in solution at neutral pH. Colour code: Cytosines are shown in green, psC in yellow, guanines in blue and thymines in magenta.

EXPERIMENTAL DISTANCE CONSTRAINTS		RMSD (Å)		AVERAGE RESIDUAL VIOLATIONS (AND RANGE)	
Total	128	Well-defined bases	0.5 ± 0.2	Sum of violation (Å)	7.2 (6.3-7.9)
Intra-residue	66	Well-defined heavy atoms	0.8 ± 0.3	Max. violation (Å)	0.78 (0.34-0.94)
Sequential	36	Backbone	1.8 ± 0.5	NOE energy (kcal/mol)	35 (24-41)
Range > 1	26	All heavy atoms	2.5 ± 0.7	Total energy (kcal/mol)	-1615 (-1490-1739)

Table 3-2. Experimental constraints and calculation statistics of **MC2_psC(9)**.

The overall structure adopted by **MC2_psC(9)** in solution at pH 7 can be described as a head-to-tail association of two DNA loops held together by six hydrogen-bonded base pairs. That is, two core C:C^+ base pairs flanked by two neutral psC:C base pairs and capped, at the outermost of the stack, by two G:G base pairs. The details of these structural features are shown in *Figure* 3-8. As it can be observed, the incorporation of psC does not alter the characteristic intercalating pattern of the motif. It is also corroborated that psC can be used as a very adequate cytidine analogue from the structural point of view, since it perfectly reproduces the hydrogen-bonding pattern and distances of the unmodified hemiprotonated C:C^+ base pairs. Contrary to what is observed in **MC2**, dynamics calculations also

confirm that the G:T:G:T minor groove tetrads are not formed, pointing to a strong capping stabilization of the structure provided by the psC:C base pairs to overcome the tetrad disruption, which is here replaced by well-defined G:G base pairs stacked at both ends of the structure.

Moreover, two thymidine residues from the loops instead of one, as **M** and **MC2** structures display, remain unstructured and exposed to the solvent since no restrictions from the formation of the tetrads are applied. *Figure* 3-8 also shows how the three-dimensional structure of this oligonucleotide affords, as it transversally occurs in more classic i-motifs as fundamental characteristic, very well-differentiated major and minor grooves.

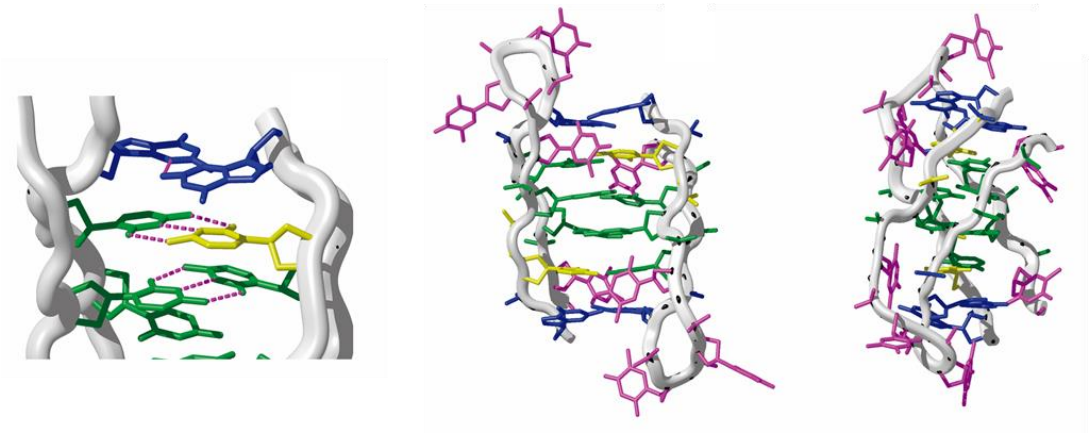


Figure 3-8. Different views and detailed features from the calculated average structure of **MC2_psC(9).** Colour code: Cytosines are shown in green, psC in yellow, guanines in blue and thymines in red.

In *Table 3-3*, average dihedral angles of the dimeric structure of **MC2_psC(9)** are shown. As commonly found, δ angles in the ring sugar are generally better defined than the rest of the torsion angles of the backbone, with order parameters close to 1. Except for psC and loop/terminal thymines, most residues exhibit dihedral angle values similar to those characteristic of right-handed duplexes (α: -50 to -60º, β: over 150º, γ: 40 to 60º, δ: ~150º, ε: 150 to 220º, ζ: -60 to -100º).[193] In fact, a variable conformation of the phosphate groups, resulting in a variable α-ζ dihedral angles had been previously reported for i-motif structures, both in XR[194] and NMR studies.[56,195] Apart from some thymine residues, basically the terminal ones, the glycosidic angle is mainly found in an *anti* conformation, with values between -90 (high anti) and -130º.

Nt	α		β		γ		δ		ε		ζ		χ	
	Avg.	OP	Avg.	OP	Avg.	OP	Avg.	OP	Avg.	OP	Avg.	OP	Avg.	OP
T1	-	-	91	0.5	85	0.4	107	1.0	-171	1.0	-	-	-56	1.0
C2	-72	1.0	173	1.0	46	1.0	115	1.0	-169	1.0	-72	1.0	-92	1.0
C3	-55	1.0	178	1.0	66	1.0	119	1.0	-172	1.0	-105	1.0	-118	1.0
G4	-68	1.0	168	1.0	46	1.0	142	1.0	-86	0.95	-78	1.0	-96	1.0
T5	-159	0.4	-176	0.9	18	0.6	126	0.9	-93	0.6	-90	0.9	-95	0.9
T6	-131	0.2	153	0.8	134	0.6	133	0.9	-136	0.81	36	0.3	-107	1.0
T7	-136	0.2	171	0.7	-141	0.5	100	0.9	-169	0.9	76	0.5	-83	0.8
C8	-62	0.6	170	0.9	72	0.5	127	1.0	-114	0.9	-93	0.8	-121	1.0
psC9	169	0.4	154	0.4	-167	0.9	137	1.0	-173	1.0	-101	0.8	-115	1.0
G10	-65	1.0	173	1.0	55	1.0	140	1.0	-100	0.8	-70	1.0	-90	1.0
T11	-94	0.8	97	0.8	-43	0.4	-1	0.9	-	-	-124	0.8	20	0.6
T1	-	-	98	0.4	-4	0.4	101	1.0	152	0.3	-	-	-53	1.0
C2	-98	0.9	177	1.0	42	1.0	114	1.0	-96	1.0	-30	0.5	-103	1.0
C3	-28	0.4	149	0.3	150	0.5	136	1.0	-166	1.0	-168	0.8	-134	1.0
G4	-63	0.7	156	0.9	32	0.9	145	1.0	-86	1.0	-81	0.9	-93	1.0
T5	-115	0.5	171	0.9	-3	0.6	124	1.0	-129	0.8	-83.5	0.9	-94	1.0
T6	167	0.3	173	0.9	167	0.6	152	1.0	-130	0.8	76	1.0	-127	0.6
T7	-125	0.4	168	0.9	177	0.5	100	0.9	-168	1.0	-4	0.1	-64	1.0
C8	-65	0.6	-175	0.7	43	0.7	122	1.0	-101	0.9	-85	0.8	-123	1.0
psC9	174	0.2	176	0.3	176	0.9	133	1.0	-171	1.0	-128	0.7	-122	1.0
G10	-56	0.7	174	1.0	39	0.8	133	1.0	-92	0.9	-89	1.0	-90	1.0
T11	-73	0.8	63	0.8	109	0.7	12	1.0	-	-	-115	0.6	-12	1.0

Table 3-3. Average dihedral angles and order parameters of the dimeric structure of **MC2_psC(9).** Average values correspond to the geometrical mean value and order parameters are use as indicators of angle definition: S=1, perfectly defined and S=0, random distribution.

Nt	Pseudorotation.		Sugar Conformation
	Phase	Ampl.	
T1	318	40	C1'-*endo*, C2'-*exo*
C2	125	42	C1'-*exo*
C3	127	45	C1'-*exo*
G4	150	46	C2'-*endo*
T5	175	25	C2'-*endo*
T6	136	39	C1'-*exo*, C2'-*endo*
T7	127	39	C1'-*exo*
C8	137	42	C1'-*exo*, C2'-*endo*
psC9	145	45	C1'-*exo*, C2'-*endo*
G10	151	37	C2'-*endo*
T11	264	45	O4'-*exo*
T1	287	40	O4'-*exo*, C1'-*endo*
C2	120	45	O4'-*endo*, C1'-*exo*
C3	143	46	C1'-*exo*, C2'-*endo*
G4	152	46	C2'-*endo*
T5	228	26	C4'-*endo*
T6	164	39	C2'-*endo*
T7	189	36	C2'-*endo*, C3'-*exo*
C8	131	43	C1'-*exo*, C2'-*endo*
psC9	139	48	C1'-*exo*, C2'-*endo*
G10	175	30	C2'-*endo*, C3'-*exo*
T11	296	54	C1'-*endo*

Table 3-4. Pseudorotation phase angle and amplitude values that allow the description of the puckering of the ribose ring in the calculated structure of **MC2_psC(9)**.

As it is presented in *Table* 3-4, the pseudorotation values obtained for the structure of **MC2_psC(9)** allow the determination of the conformation of the sugar ring for all the residues. Notably, as a general trend, most of the sugars adopt South-East conformations. Although C2'-*endo* puckering is the most common, C1'-*exo* and O4'-*endo* conformations are also observed. The exceptions to this generality include thymine residues at the 5' and 3'-end of the sequence and some residues involved in the loops of the structure. Sugar conformations differ from those generally found for cytosine residues in i-motif structures (C3'-endo and C4'-exo).[56] This may be related to the fact that, in this case, the core of the structure is not formed by a uniform stack of hemiprotonated $C:C^+$ base pairs.

3.3.1.3. A HEAD-TO-HEAD SPECIES EXISTS AS A MINOR FORM AT ACIDIC pH

The extra imino signals at ~ 15 ppm in the 1D ^{1}H-NMR spectra at pH 6 and pH 5 suggest the formation of alternative i-motif structures for the **MC2_psC(9)** sequence under acidic conditions. The characterization of the different species was performed with NOESY and TOCSY experiments at pH 5. Although the better resolution of the signals at this pH facilitated their assignment, the complete assignment of the different species could not be achieved due to the overlapping between key signals.

The initial information obtained from the exchangeable proton region of the NOESY spectra at acidic pH matches with the coexistence of head-to-head and head-to-tail dimeric species in equilibrium. The three imino signals characteristically shifted downfield of the $C:C^+$ base pairs could be assigned to the formation of three $C:C^+$ base pairs between symmetry-related (or equivalent) residues, and to one $C:C^+$ base-pair between no equivalent cytosine. The imino signal at 15.53 ppm shows NOE cross-peaks with six amino protons, corresponding to overlapped imino protons of one $C:C^+$ base pair between non-equivalent residues ($C2/C8:C8/C2^+$, connected with two pairs of amino protons) and one $C:C^+$ base pair between equivalent residues ($C3:C3^+$, connected with one amino protons pair). Each of the other two hemiprotonated cytidine imino signals (15.62 and 15.29 ppm) show two cross-peaks with amino protons, confirming the $C:C^+$ pairing between equivalent residues ($C2:C2^+$ and $C8:C8^+$). Unfortunately, unambiguous assignment of the C2/C8 residues could not be provided due to the complete overlapping between sequential cross-peaks. This information supports the loss of symmetry that takes place upon formation of the head-to-head dimer.

Two imino signals presumably corresponding to the psC residues in each of the dimeric species are found in the diagonal of the NOESY spectra. Unfortunately, we could not observe any cross-peak with amino protons, preventing their assignment to a particular species.

Regarding the stacking order of the residues, although the unambiguous identification of C2 and C8 could not be assessed, H42-H2'/H2"C2/C8 cross-peaks confirm that these residues are located in the center of the stem, facing their 3' sides, as found for the head-to-tail topology. Stacking order is thus the same for both orientations.

Overall, we concluded that the major species at neutral and acidic pH is the head-to-tail dimer, since the cross-peak pattern from the NOESY experiments remains essentially unaltered from pH 7 to pH 4. However, we can confirm the presence of the head-to-head dimer at acidic pH, based on the established connections between exchangeable protons of C:C$^+$ base pairs between equivalent residues.

3.3.2. MC2_psC(8)

3.3.2.1. **MC2_psC(8)** FOLDS INTO A MAJOR HEAD-TO-HEAD SPECIES AT ACIDIC pH

Taking the structure of **MC2** as a model, substitution of the cytidine residue in position 8 by psC was expected to render an i-motif structure with psC:C (head-to-tail topology) or psC:psC (head-to-head topology) base pairs at the core of the structure. However, this hypothesis could not be confirmed under any experimental conditions. As it was outlined by 1D ^{1}H-NMR experiments, i-motif formation for the sequence **MC2_psC(8)** is only observed under acidic conditions (*Figure* 3-3). Contrary to the single structures observed for both **MC2** and **MC2_psC(9)**, the well-dispersed exchangeable signals of the NOESY spectra of **MC2_psC(8)** at pH 5, suggest the presence of at least three different species in equilibrium (up to eight cytidine and two psC residues could be identified belonging to structured or partially structured species).

The best-defined set of connected spin systems could be assigned to a dimeric i-motif structure with head-to-head topology and thus defined as major species. Displayed in *Figure* 3-9, four imino signals at around 15 ppm, each connected to a pair of amino protons (H3$^+$-H41/H42), assessed the formation of four hemiprotoanted base pairs between equivalent residues. This information led us to assume that the major species consists of a head-to-head dimer with three C:C$^+$ and one psC:psC$^+$ base pair (*Figure* 3-10*)*. The assignment of one of these imino signals to a hemiprotoanted base pair between psC residues was assessed by the lack of individual H41/H42-H5 cross-peaks. Interestingly, the downshift of the imino signal of the psC:psC$^+$ base pair compared to the neutral psC:C allows to confirm its protonation.

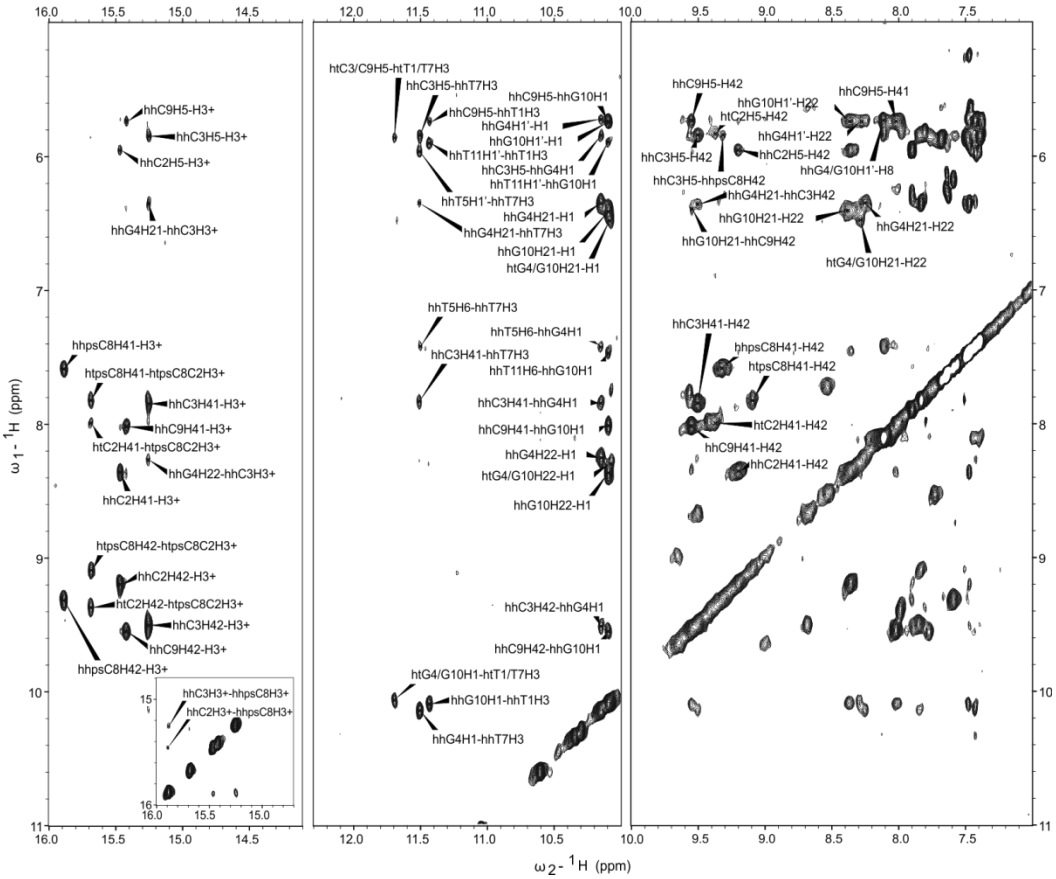


Figure 3-9. Imino and amino exchangeable protons region of NOESY spectrum (150 ms) of **MC2_psC(8)** . H_2O/D_2O 90:10, 25 mM phosphate buffer pH 5, 100 mM NaCl, T = 5 °C, [oligonucleotide] = 0.8 mM. hh labels correspond to the signals of the head-to head dimer. ht labels correspond to the signals of the head-to tail dimer.

The stacking order of the hemiprotonated base pairs was determined through observation of major groove H42-H2'/H2" contacts between psC8 and C2 residues, and through imino-imino cross-peaks that clearly connect the C3-psC8-C2 path. Stacking H41/H42C-H1G and H1G-MeT cross-peaks allowed to complete the stacking order, which was assigned to T11-G10-C9-C2-psC8-C3-G4-T5, analogous to **MC2** and **MC2_psC(9)**.

Notably, characteristic imino-imino cross-peaks between guanine and thymine residues were used to assess G:T:G:T minor groove tetrads formation (T1-G10-G10-T1 and T7-G4-G4-T7). The observation of these tetrads in **MC2_psC(8)** but not in **MC2_psC(9)**, suggests that their formation relies strongly on the presence of adjacent hemiprotonated base pairs.

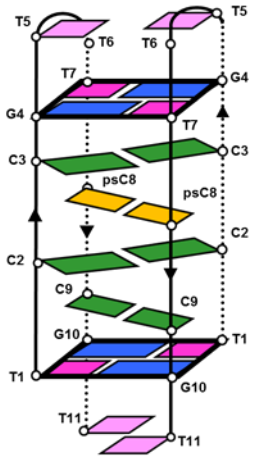


Figure 3-10. Schematic representation of the head-to-head major structure of **MC2_psC(8)** at pH 5.

Based on the number of cytidine residues spin systems, two partially structured minor species were found coexisting with the major head-to-head structure. The high overlapping of the signals and the lack of some key exchangeable proton region cross-peaks impeded the complete characterization of these minor species. Remarkably, the presence of extra psC residues involved in psC:C^+ base pairing and G:T:G:T minor groove tetrads cross-peaks were identified. However, no further insights could be provided. As a hypothesis, the structures of the minor species of **MC2_psC(8)** at acidic pH were assigned to head-to-tail and head-to-head dimers with different stacking order than the major species.

These results provide important information regarding the preference of forming hemiprotoanted base pairs rather than including neutral psC:C pairs in the center of the i-motif structure. psC:C base pairs clearly destabilizes the structure unless placed at outmost positions of the stack. Moreover, probably due to the lower pK_a values of psC, the topology of the structure appears to be determined by the maximum number of C:C^+ (head-to-head conformation) base pairs rather than psC:C^+ (head-to-tail topology).

3.3.3. M_psC(7)

3.3.3.1. **M_psC(7)** EXHIBITS A pH-DEPENDENT MIXTURE OF CONFORMATIONS AT ACIDIC pH

As in the case of **MC2_psC(8),** NMR spectra indicate that **M_psC(7)** is only folded under acidic conditions (*Figure* 3-3). Compared to its 11-mer analogues, the relative populations of **M_psC(7)** species (represented in *Figure* 3-11) are different at pH values ranging from pH 5 to pH 4. This fact is probably due to the different stability of the psC-containing base pairs depending on their protonation state.

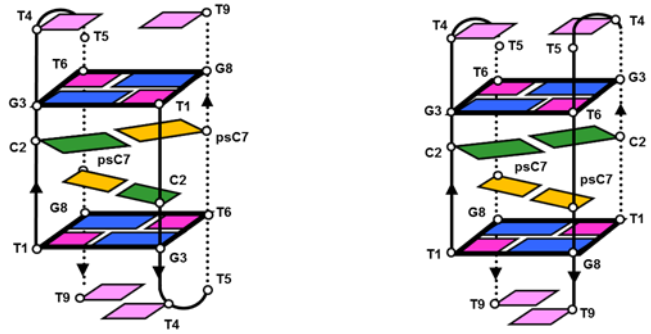


Figure 3-11. Schematic representations of the head-to-tail major structure of **M_psC(7)** at pH 5 (left) and the head-to-head major structure of **M_psC(7)** at pH 4 (right).

At pH 5, NOESY spectra are consistent with the formation of these two dimers of different topology (*Figure* 3-12). The two imino signals observed at ~ 15 ppm present different cross-peak patterns, with one of them (15.27 ppm) connected to a single pair of amino protons, and the other (14.70 ppm) connected to four amino signals. This information is compatible with the formation of one species formed by hemiprotonated base pairs between non-equivalent residues, and another species with at least one base pair between equivalent residues. One of the amino pairs connected to the imino proton signal at 14.70 ppm was assigned to a psC residue as it did not present H41/H42-H5 cross-peaks. Thus, for **M_psC(7)** at pH 5, a head-to-tail dimer, containing two psC:C$^+$ base pairs, and a head-to-head dimer, with one C:C$^+$ base pair, were identified. The potential formation of a psC:psC$^+$ base pair in the head-to-head species could not be assessed under these experimental conditions.

The formation of G:T:G:T minor groove tetrads was determined for both species with the assignment of representative H3T-H1G and H1'G-H1G cross-peaks. Moreover, the stacking order T9-G8-psC7-C2-G3-T4 was established on the basis of stacking interactions (H41/H42C(psC)-H1G and H1G-MeT) between C(psC)-G-T residues, confirming the high preference of C:C$^+$ base pairs to be perfectly stacked above/below the tetrads.

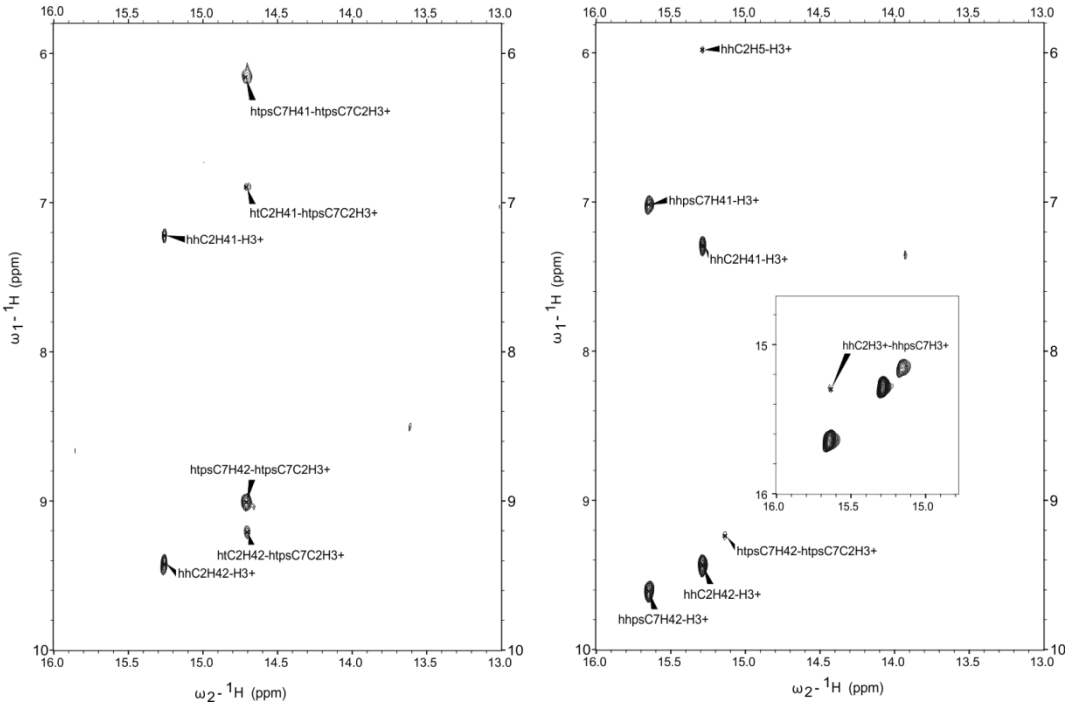


Figure 3-12. Imino exchangeable protons region of NOESY spectra (150 ms) of **M_psC(7)** at pH 5 (left) and pH 4 (right). H_2O/D_2O 90:10, 25 mM phosphate buffer, 100 mM NaCl, T = 5 °C, [oligonucleotide] = 0.8 mM. hh labels correspond to the signals of the head-to head dimer. ht labels correspond to the signals of the head-to tail dimer.

Interestingly, increasing the acidity of the solution to pH 4, resulted in an apparent shift of the equilibrium between the head-to-tail and the head-to-head topologies. If at pH 5 the head-to-tail conformation seemed to be slightly favored, the head-to-head species is clearly predominant at pH 4. This information is supported by the appearance of a new imino proton signal at 15.66 ppm assigned to a psC residue involved in the formation of a base pair between equivalent residues (see *Figure* 3-12). An imino-imino cross-peak undoubtedly connects this psC:psC$^+$ base pair with the C2:C2$^+$ base pair already observed at pH 5. Also, the imino-amino cross-peaks that defined the psC:C$^+$ base pairs of the head-to-tail topology at pH 5, are almost undetectable at pH 4, although amino-amino cross-peaks of the psC residue remain unaltered.

The fact that the replacement of cytidine by psC in position 7 in **M**, which would be analogous to substitution in position 9 in **MC2**, does not afford stable i-motif structures at neutral pH, can be interpreted as the key importance of having at least some hemiprotonoated base pairs in the structural core of the motif. This suggests that i-motifs cannot be stabilized only with neutral psC:C base pairs.

3.4. INTRODUCING psC IN THE HUMAN TELOMERIC i-MOTIF

As it was suggested in the introductory section of this chapter, although **M** and **MC2** are very convenient and interesting sequences for studying the effect of a neutral analogue of protonated cytidine such a psC in i-motif structures, it is interesting to compare the previous results with other well-known i-motif forming sequences. To expand the scope of our results to a more general model of i-motif structure, we explored the effect of psC incorporation in the C-rich strand of the human telomeric sequence, named **HT0,** d(CCCTAACCCTAACCCTAACCC). As shown in *Figure* 3-13 the structure of **HT0** has been described as a unimolecular i-motif formed by six C:C$^+$ base pairs with 5'-E topology and TAA loops.[196] Three psC-containing sequences were synthesized. **HT_psC(1)** contains a single psC residue ideally designed to occupy the outermost and less impeded position of the C:C$^+$ stack. **HT_psC(1,7)** and **HT_psC(2,8)** were designed to both contain two psC residues as a mirror to **MC2_psC(9)** and **MC2_psC(8)**. While **HT_psC(1,7)** should fold into an i-motif structure with two psC:C base pairs capping the ends of the C:C$^+$ stack, **HT_psC(2,8)** is meant to form an i-motif structure with psC:C base pairs at its central positions (*Figure* 3-13). See *Section* 7.2.1 in the Syntheses chapter for experimental details on the synthesis of these sequences.

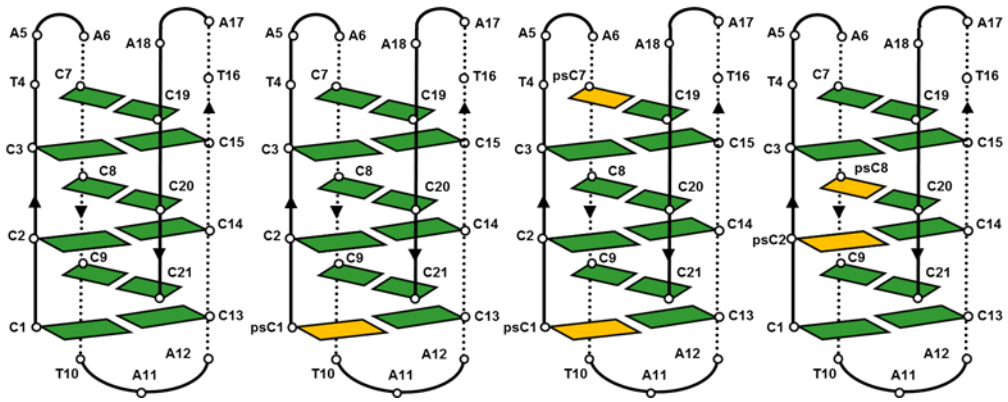


Figure 3-13. Schematic representations of **HT0** (left), **HT_psC(1)** (center left), **HT_psC(1,7)** (center right) and **HT_psC(2,8)** (right) structures.

3.4.1. FORMATION AND STABILITY

Followed by 1D ^{1}H-NMR experiments (see *Figure* 3-14), i-motif formation at neutral pH was assessed from the characteristic C:C$^+$ imino signals observed at around 15 ppm for all the sequences. Interestingly, in the case of **HT_psC(2,8)** these signals are not observed, in agreement with our previous observation that neutral psC:C base pairs are not allowed in the central positions of the motif.

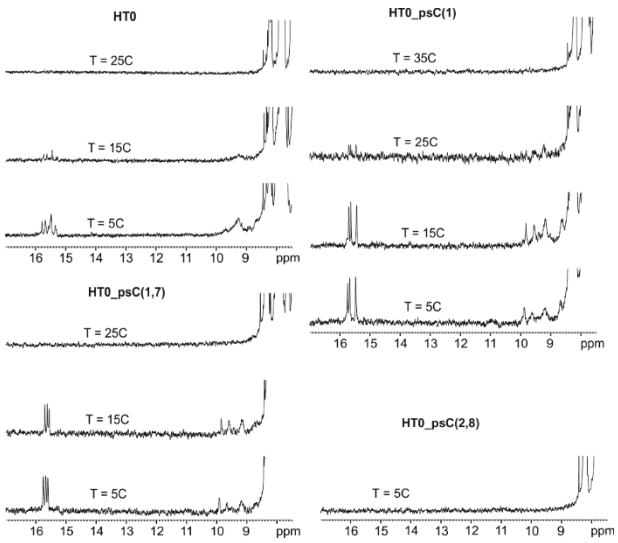


Figure 3-14. ^{1}H-NMR spectra melting series. H_2O/D_2O 90:10. [oligonucleotide] = 100 µM, 10 mM phosphate buffer, pH 7.

CD spectra recorded at different temperature at pH 6.8 and pH 5.5 are shown in *Figure* 3-15. As a first remarkable observation, one must observe the shift in wavelength value of the maximum band in each spectrum belonging to structured species. As for mini i-motif structures the maximum band is found at 265 – 270 nm,[191] telomeric sequences show here more classic i-motif CD spectra with maximum bands between 290 – 295 nm. The series at pH 6.8 (*Figure* 3-15, left) shows a very similar behaviour of **HT0** and **HT_psC(1)**. On the other hand, at acidic pH, the CD spectra of **HT_psC(1)** shifts to the random coil spectra at slightly lower temperature values than the native sequence. The notable low stability of **HT_psC(2,8)** is very noticeable in *Figure* 3-15, which shows no sign of structuration for this sequence at any temperature at pH 6.8. Thus, it is confirmed the high destabilization provoked by pseudoisocytidine residues incorporated to central positions of the hemiprotonated C:C$^+$ stack. Finally, **HT_psC(1,7)** exhibits a similar stability to that of **HT_psC(2,8)** at acidic pH, suggesting that the protonation of psC affects differently to the stability of the structure depending on the position in which is located. Altogether, these results confirm the position-dependency of psC as a stabilizing/destabilizing agent of i-motif structures.

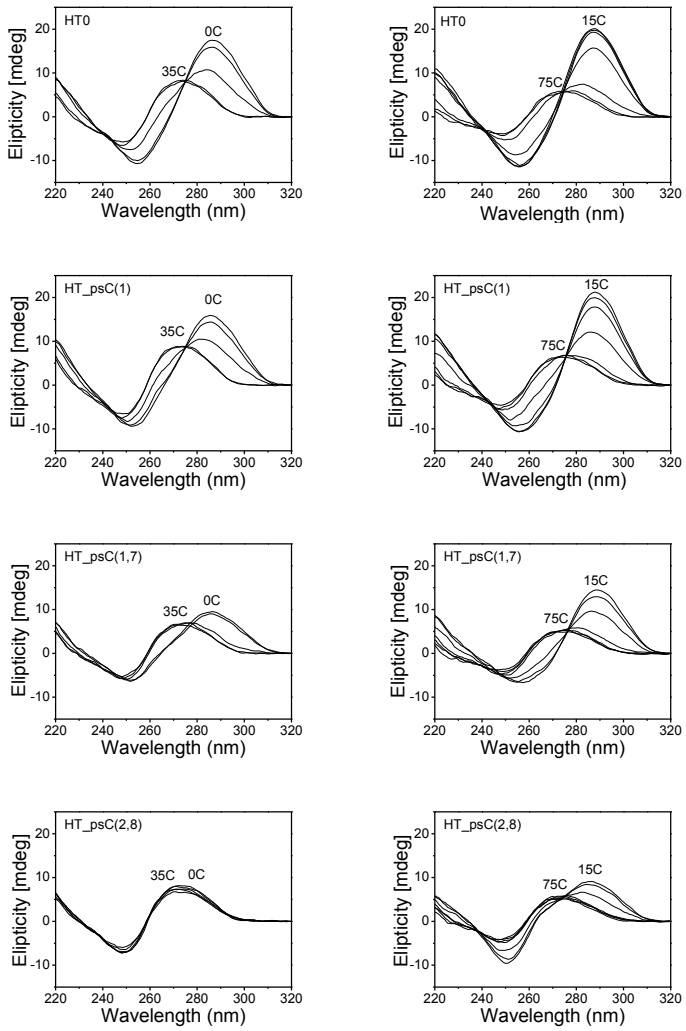


Figure 3-15. CD spectra superposition of the psC-containing telomeric sequences at different temperatures. Left: pH 6.8 and acquisition at 0, 5, 15, 25 and 35 ºC. Right: pH 5.5 and acquisition at 15, 25, 35, 45, 55, 65 and 75 ºC. [oligonucleotide] = 20 μM, 25 mM cacodylate buffer solution.

The thermal stability of the structures was studied by 1D ^{1}H-NMR (*Figure* 3-14 and *Figure* 3-17) and CD experiments. The thermodynamic behavior of the psC-containing telomeric sequences is notably different. These differences, as the following discussion enlightens, matches the position-dependency of psC found for the previously presented dimeric i-motifs.

CD melting experiments were performed at slighter acidic (6.8) and acidic pH (5.5 – 6). The quantitative results of these experiments are presented in *Table* 3-5 and the respective melting curves are shown in *Figure* 3-16. Since the obtained T_m values at pH 7 are very low, error may be significantly higher than those shown in *Table* 3-5, arising from the melting curves fitting.

NAME	SEQUENCE	T_m @pH5.5	T_m @pH6.8
HT0	d(CCCTAACCCTAACCCTAACCC)	45.0 ± 0.5	14.0 ± 0.5
HT_psC(1)	*d(**psC**CCTAACCCTAACCCTAACCC)*	43.0 ± 0.5	14.2 ± 0.5
HT_psC(1,7)	*d(**psC**CCTAA**psC**CCTAACCCTAACCC)*	36.9 ± 0.5	11.3 ± 0.5
HT_psC(2,8)	d(C***psC***CTAAC***psC***CTAACCCTAACCC)	35.4 ± 0.5	-

Table 3-5. Sequences and stability data against temperature (ºC) obtained from CD experiments. Experimental conditions: [oligonucleotide] = 20 μM, 25 mM of buffer solution (cacodylate (pH 5.5) or phosphate (pH 6.8)).

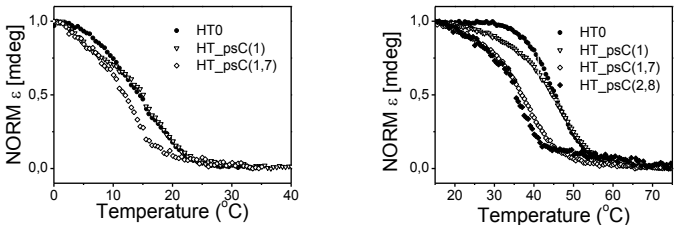


Figure 3-16 Melting curves of the HT0-derivative sequences at pH 6.8 (left) and 5.5 (right). **HT_psC(2,8)** remains unfolded at any temperature at pH 6.8. [oligonucleotide] = 20 μM, 25 mM cacodylate buffer solution.

The conclusion that can be extracted from these experiments is that psC provokes a general destabilization of the structures. However, while the destabilizing effect in **HT_psC(2,8)** is dramatic, for the other psC-containing telomeric sequences is substantially lower. In fact, the single introduction of psC in **HT_psC(1)**, located at the end of the C:C^+ stack, although affording a slightly lower T_m value than the native sequence at acidic pH, appears to be responsible for the enhanced stability of the structure at neutral pH, especially noticeable from 1D NMR experiments shown in *Figure* 3-17. Analogously, the NMR melting series of **HT_psC(1,7)** at pH 7 apparently show a slight stabilization of this structure compared to the native **HT0** based on the intensity of the imino protons at 15 ppm that is not, however, supported by the CD melting experiments (*Table* 3-5).

Figure 3-17 shows the NMR melting series of the HT0-derivative sequences at pH 6. The higher stability of the native HT0 sequence is evident, as expected from the reported pH_T values (6.2 - 6.5)[96,197] of different constructions of the sequence. The tendency observed at neutral pH is maintained, with **HT0** and **HT_psC(1)** showing the best behavior among all the sequences.

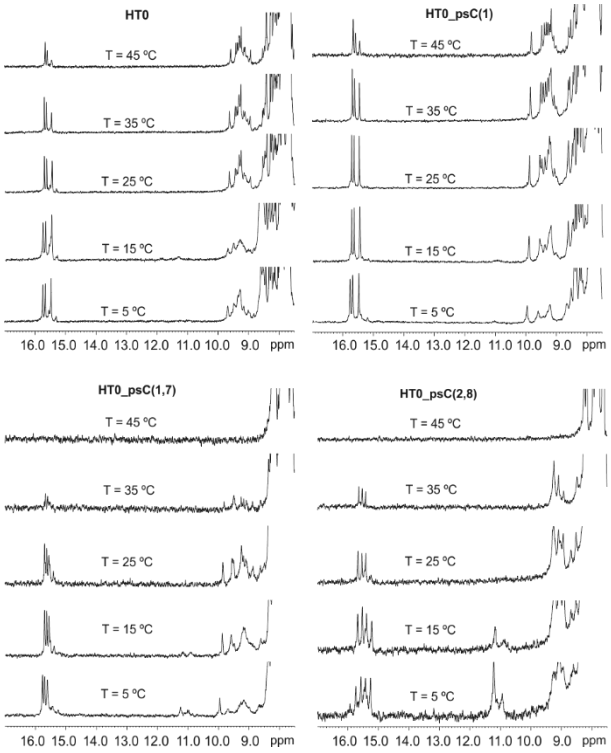


Figure 3-17. ^{1}H-NMR spectra melting series. H_2O/D_2O 90:10. [oligonucleotide] = 100 µM, 10 mM phosphate buffer, pH 6.

3.5. CONCLUSIONS

Pseudoisocytidine was successfully introduced in several oligonucleotide sequences by common SPOS methods and its compatibility with i-motif formation was assessed by different characterization techniques. The i-motif structures were characterized on the basis of NMR data at different pH conditions. In the favorable case of **MC2_psC(9),** the detailed three-dimensional structure was determined by restrained molecular dynamics methods on the basis of the experimental NMR constraints.

The effect of psC on the stability of the i-motif structures strongly depends on the position in which the cytidine replacement takes place. While neutral psC:C base pairs located at the ends of the C:C$^+$ stack enhance the stability of the motif, affording relatively stable structures at neutral pH, incorporation of psC in central positions of the structure prevents i-motif formation. Compared to other chemical modifications, the psC stabilization degree is comparable to that of 5-methylcytosine,[97] LNA[198] and some sugar modifications.[105] Analogously, its position-dependent destabilizing effect is similar to the ones provoked by 5-hydroxymethylcytosine[97] or UNA.[199]

The i-motif structures designed to contain psC residues in the central positions of their stack are less stable than they unmodified analogues and involve the formation of hemiprotonated psC:C^+ or psC:psC^+ base pairs. Therefore, are only observable under acidic conditions.

Based on these observations, the importance of protonation in i-motif structures pointed out. The presence of positive charges is indispensable for the formation of the motif, possibly in order to overcome electrostatic repulsions triggered by close sugar-phosphate backbones. Moreover, the near presence of positive charges also appears to be essential for the formation of G:T:G:T minor groove tetrads in i-motif structures, suggesting that their stabilizing effect may be provoked by cation-pi interactions with the hemiprotonated base pairs.

The overall results and conclusions describe in this chapter have been published.[200]

4. MINOR GROOVE TEDRADS IN i-MOTIFS. PART I: DIMERIC STRUCTURES

4.1. BACKGROUND AND OBJECTIVES

Parallel to the attempt presented in the previous chapter, the stabilization of i-motif structures at neutral pH was also challenged by exploring if the formation of these structures can be enhanced by modulating the possible capping interactions found in these motifs.

As it has been widely introduced, slipped minor groove G:T:G:T tetrads have been found as capping elements in i-motif structures of CENP-B box derived sequences.[86] The stabilization conferred by these tetrads at the ends of the C:C$^+$ stack is so remarkable, that they have allowed to observe the formation of i-motif structures consisting of only two hemiprotonated C:C$^+$ base pairs even at neutral pH.[191] The observation of these mini i-motifs, with no contiguous cytosines in the sequence, widens the scenario where i-motif structures can be found, probably not limited to C-rich sequences.

Consequently, we found interesting to explore the compatibility of other minor groove tetrads, found in different structural contexts, with i-motif structures (see *Figure* 4-1 and *Figure* 4-2). The formation of different minor groove tetrads has been observed, for instance, as additional stabilizing contacts in crystal structures of A- and B-DNA duplexes. In the context of B-DNA, slipped G:C:G:C minor groove tetrads result from the interaction of stacked duplexes through the minor groove side of G:C WC base pairs.[201–203] In A-DNA crystals, in contrast, two different types of minor groove tetrads have been reported: direct G:C:G:C[204] and direct G:C:C:G tetrad.[205] In both cases, interacting WC base pairs are not displaced but found directly facing guanine and cytosine residues. Interestingly, a greater variety of minor groove tetrads were later found in the context of bi-loop structures. These peculiar constructions can be described as highly compact dimeric structures stabilized by the formation of two stacked minor groove tetrads. This type of motif was first described for the linear heptamer d(GCATGCT) and the cyclic octamer d<pCATTCATT>, stabilized by the formation of direct G:C:G:C and direct A:T:A:T minor groove tetrads, respectively.[206,207] The term “bi-loop” has been proposed for this family of structures, which has been extensively studied in our research group. In this context, a number of different types of minor groove tetrads have been reported, such as direct G:C:G:C,[208,209] direct G:C:A:T,[87] slipped G:C:G:C,[210] direct C:G:G:C[211] and slipped G:C:G:T and G:T:G:T tetrads.[212]

Except for A:T:A:T and G:C:A:T, the formation of minor groove tetrads results in the formation of two additional hydrogen bonds between two aligned base pairs (see *Figure* 4-1). In direct G:C:G:C and C:G:G:C tetrads, both cytosine and guanine bases are involved in these extra bonds: two H21(G)-O2(C) and one H21(Ga)-O2(Cb) and one H21(Gb)-N3(Ga), respectively. On the contrary, in all slipped tetrads, the extra bonds are formed exclusively between guanine residues: H21(Ga)-N3(Gb) and H21(Gb)-N3(Ga). Regarding A:T:A:T and G:C:A:T no extra hydrogen bonds can be formed when interacting through the minor groove and stabilization requires cation coordination, probably to the O2 oxygen atom of thymine residues (*Figure* 4-2).

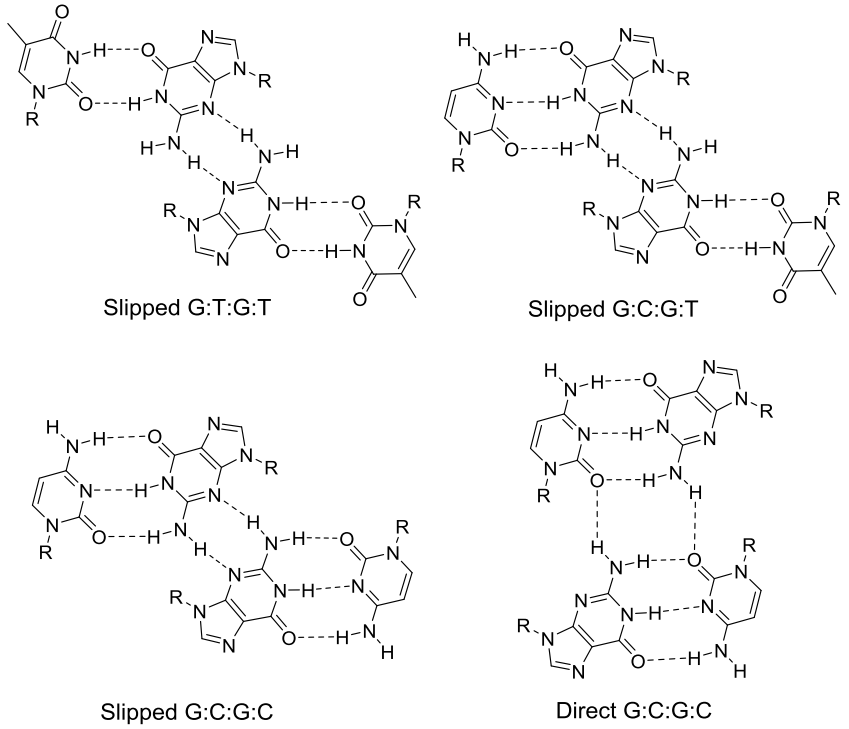


Figure 4-1. Minor groove tetrads stabilized by extra minor groove hydrogen bonds.

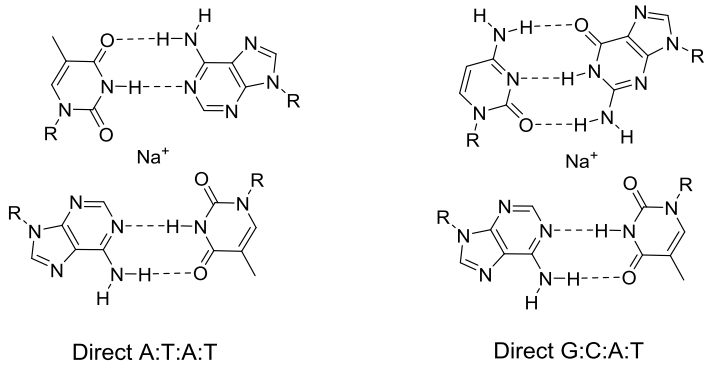


Figure 4-2. Minor groove tetrads stabilized by cation coordination.

A common feature shared by all these minor groove tetrads is that the base pairs involved in their formation exhibit a relative inclination of approximately 30-40º. This contrast with major groove tetrads, usually found within G-quadruplex scaffolds, which are planar. A great variety of major groove tetrads

have also been reported, including direct G:C:G:C,[213–215] A:T:A:T[216] and G:T:G:T[217] and slipped G:C:G:C[218] and A:T:A:T tetrads.[219]

In the i-motif structures were slipped G:T:G:T minor groove tetrads have been found so far,[191,220] these type of tetrads fit perfectly at the end of the C:C$^+$ stack providing favourable stacking interactions to the overall stability of the structures (*Figure* 4-3). On the contrary, major groove tetrads seem to be not compatible with the formation of i-motif constructions due to backbone restrictions.

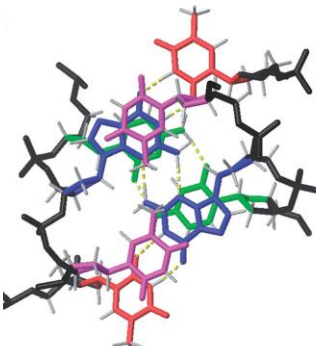


Figure 4-3. Details of the stacking interaction between C:C$^+$ base pairs with slipped G:T:G:T minor groove tetrads and capping thymine[76]. Color code: cytosines-green, guanine-blue, bonded thymine-red, loop thymine-magenta.

With the aim to explore the compatibility of i-motif structures with G:T:G:T and other minor groove tetrads, we have studied different linear and cyclic oligonucleotides derived from the **M** sequence d(TCGTTTCGT)[76]. In particular, we will study sequences able to render mini i-motif structures capped by the other reported slipped tetrads: G:C:G:T[212] and G:C:G:C[210,219,221]. In addition, we have included a sequence that may form A:T:A:T minor groove tetrads. Slipped A:T:A:T tetrads have only been reported as major groove tetrads.[219] We want to test, if the stacking between adenine and hemiprotonated cytosine residues may drive the formation of slipped tetrads in this case. In hypothesized slipped A:T:A:T tetrads, no additional hydrogen bonds can be formed, although a cation could be coordinated to the N3 nitrogen atoms of adenine residues. By exploring multiple sequences, we will also know whether minor grove tetrads always adopt the slipped conformation or, on the contrary, direct minor groove tetrads are also possible in the i-motif context.

4.1.1. SEQUENCES

The oligonucleotides studied in this chapter, which are presented in *Table* 4-1, arise from strategic nucleotide substitutions performed on the original **M** sequence. As described in previous chapter, this sequence folds into dimeric mini i-motif structures, capped at both ends by slipped G:T:G:T minor groove

tetrads. As previously mentioned, an equilibrium between the two possible strand orientations, head-to head and head-to-tail, (*Figure* 4-4) was observed in solution.

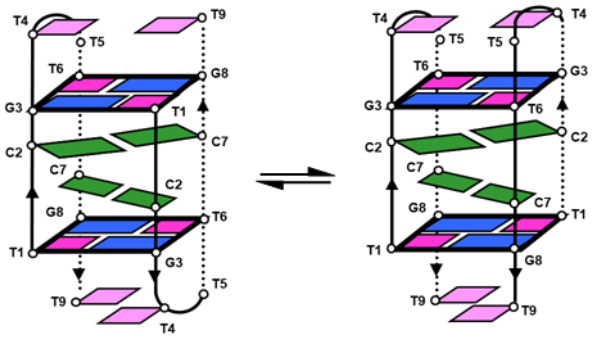


Figure 4-4. Head-to-tail (left) and head-to-head (right) topologies of the **M_2** dimer.

As in previous studies of the sequence **M**, both linear and cyclic oligonucleotides have been considered. In addition to the higher stability exhibited by **cM** compared to the linear sequence **M**,[76] the use of cyclic analogues may facilitate the NMR assignment and structural characterization of the sequences for several reasons: First, the number of spin systems is reduced due to the higher symmetry and, second, the number of alternative structures is limited by the conformation constraint induced by the cyclization.

Each sequence will be particularly introduced in the following sections of this chapter, together with a proper discussion of the nature of the tetrads that we expect to observe, and their influence on the formation and stability of the structures.

Some of the oligonucleotides utilized in this chapter were synthetized by standard SPOS. The results of the syntheses are summarized in *Section* 7.2.2. The method for the synthesis of the cyclic oligonucleotides is described in *Section* 7.1.2.

NAME	SEQUENCE	EXPECTED TETRADS
M	d(TCGTTTCGT)	G:T:G:T
cM	d<pTCGTTTCGTT>	G:T:G:T
L	d(TCGTT**C**CGT)	2 G:C:G:T or G:T:G:T & G:C:G:C
cL	d<pTCGTT**C**CGTT>	2 G:C:G:T or G:T:G:T & G:C:G:C
N	d(**C**CGTT**C**CGT)	G:C:G:C
cN	d<p**C**CGTT**C**CGTT>	G:C:G:C
K	d(TC**A**TTTC**A**T)	A:T:A:T

Table 4-1. Oligonucleotide sequences studied in this chapter and possible MGT that may be formed upon mini i-motif formation.

4.2. EVALUATION OF I-MOTIF FORMATION AND pH-DEPENDENCE

To evaluate the formation of i-motif structures, NMR and CD techniques s will be used in order to determine and compare the stability of the different oligonucleotides under different experimental conditions, with special focus on the dependence of the structures upon pH modifications.

4.2.1. L-RELATED SEQUENCES

4.2.1.1. **L,** d(TCGTTCCGT)

The design of this sequence includes a single nucleotide replacement in the sixth position of **M**, including a cytidine residue substituting a thymidine. This substitution, based on the solution structure of the $\mathbf{M_2}$ dimers, would have an important effect on the nature of the minor groove tetrads, as two of the G:T mismatched base pairs are changed to WC G:C base pairs. Interestingly, a single nucleotide substitution in **L**, provokes a loss of symmetry within the sequence that would be translated into the formation of different type of minor groove tetrads depending on the strand orientation upon dimerization. More specifically, a head-to-tail orientation would imply the formation of two identical G:C:G:T tetrads, while in a head-to-head orientation one G:T:G:T and one G:C:G:C minor groove tetrad would be expected (see *Figure* 4-5).

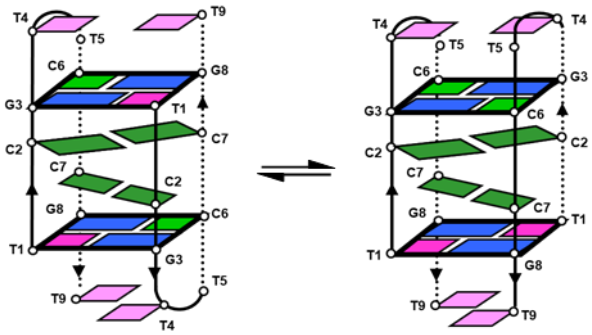


Figure 4-5. Expected head-to-tail (left) and head-to-head (right) topologies for the $\mathbf{L_2}$ dimer.

1D ^{1}H-NMR spectra recorded at pH 7 unequivocally confirm the formation of a well-defined i-motif-like structure for the **L** oligonucleotide (*Figure* 4-6). Moreover, although further structural studies are required, the reduced number of exchangeable protons signals and their clean dispersion suggest the formation of a unique major species at neutral pH. As a first insight on the stability of this i-motif structure, i-motif signals are observable up to 25 ºC under these experimental conditions and, most importantly, all exchangeable protons signals seem to disappear concomitantly, indicating an interdependence of all the structural features in order to maintain the motif folded. Besides the characteristic imino signal of

hemiprotonated C:C$^+$ base pairs found above 15 ppm, an intense signal at around 14 ppm clearly indicates the formation of WC G:C base pairs. Moving upfield in the spectra, a less intense signal between 12 and 13 ppm might belong to hybridized imino protons of guanine or thymine residues. Finally, the two intense signals right below 11 ppm most likely belong to amino protons of cytosine residues.

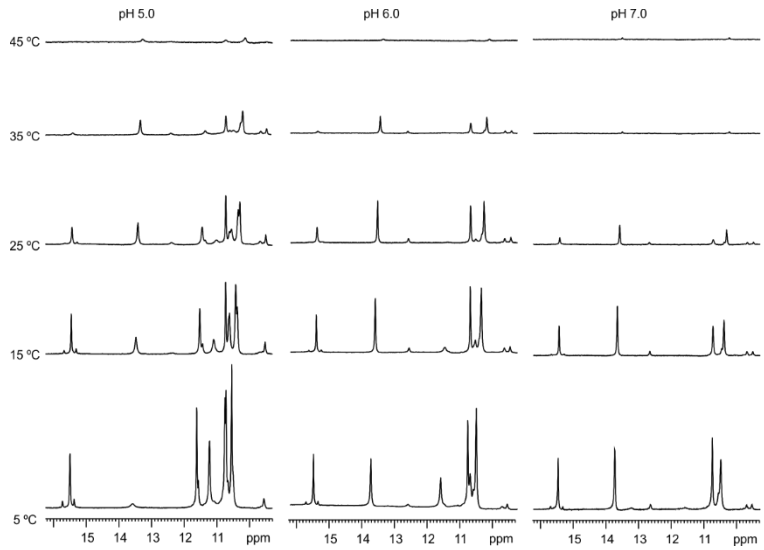


Figure 4-6.1D ^{1}H- NMR melting series of **L** at different pH values. H_2O/D_2O 90:10, 10 mM phosphate buffer solution and 100 mM NaCl. [oligonucleotide] = 1.8 mM

At pH 6, the previously described signals remain unaltered and some extra signals appear. Stability-wise, exchangeable protons are now observable up to 35 ºC. Apart from some extra amino signals at around 11 ppm, most interestingly, the new signal between 11 and 12 ppm strongly suggests the formation of G:T mismatches which, together with the C:C$^+$ and the WC G:C base pairs, completes the set of base pairs expected for the mini i-motif of **L** independently of its topology.

Finally, at pH 5, some residual signals are observable up to 45 ºC, indicating as expected a higher stability of the structures at acidic pH. Under these conditions, the intensity of the signals between 10 and 12 ppm is notably increased also appearing some new signals. On the other hand, the imino signal of the guanine residue involved in the G:C base pair becomes less intense, probably due to the partial protonation of the cytosine residues. Moreover, low intensity signals above 15 ppm that were residual at pH 6 and 7, become more evident at pH 5, suggesting the formation of an alternative minor i-motif species at acidic pH that will have to be evaluated by 2D NMR experiments.

4.2.1.2. **cL**, d<pTCGTTCCGTT>

As in the case of the **L** sequence, for **cL** also two possible structures, stabilized by either two G:C:G:T tetrads or one G:C:G:C and one G:T:G:T tetrads, are expected based on the relative rotation of one cycle respect to the other, as it is shown in *Figure* 4-7.

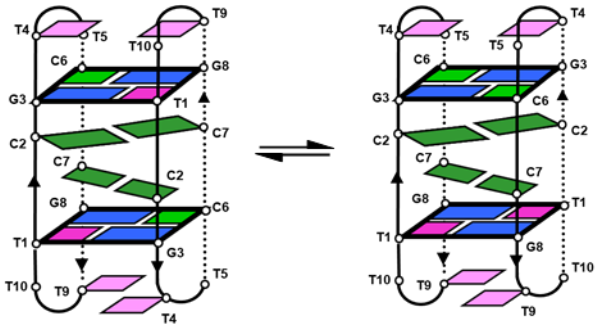


Figure 4-7. Possible topologies of the **cL$_2$** dimer.

Unfortunately, due to the low oligonucleotide concentration of the sample provoked by the low yield obtained in the synthesis of cyclic oligonucleotides (*Section* 7.2.2), the NMR spectra recorded at pH 7 are of low quality and include signals corresponding to unstructured species and probably intramolecular dumbbell-like associations. Therefore, in order to focus on major i-motif structures, the study of this sequence was performed under acidic experimental conditions.

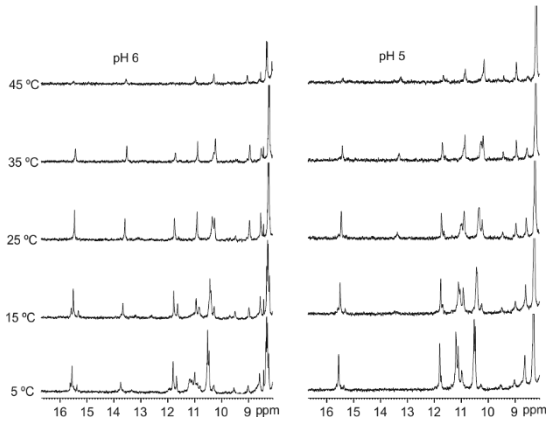


Figure 4-8. 1D ^{1}H-NMR melting series of **cL** at different pH values. H_2O/D_2O 90:10, 10 mM phosphate buffer solution and 100 mM NaCl. [oligonucleotide] = 0.6 mM

NMR spectra presented in *Figure* 4-8 for **cL**, show similar exchangeable protons signals pattern than that found for its linear analogue. Signals consistent with the formation of $C:C^+$ (above 15 ppm), WC G:C (between 13 and 14 ppm) and G:T wobble base pairs (around 12 ppm) coexist at pH 6 and exhibit the same behavior upon temperature changes. Moreover, the disappearance of the imino signal corresponding to the G:C base pair at pH 5 is also observed. Therefore, we can conclude that the i-motif structures adopted by the linear and cyclic oligonucleotides are very similar and that their folding/unfolding dynamics are comparable. In terms of stability, as expected for cyclic oligonucleotides, **cL** remains folded at higher temperature values than **L**.

4.2.2. N-RELATED SEQUENCES

4.2.2.1. **N**, d(CCGTTCCGT)

By substituting the thymidine residues in positions one and six of **M** by cytidines, we expect the dimeric i-motif formed by the **N** sequence to be stabilized by two G:C:G:C minor groove tetrads at both ends regardless the topology adopted by the $\mathbf{N_2}$ dimer, as it is shown in *Figure* 4-9.

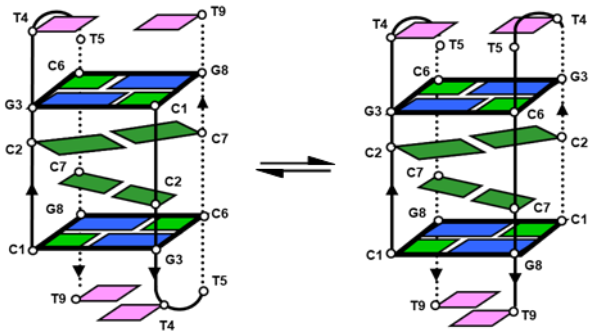


Figure 4-9. Expected head-to-tail (left) and head-to-head (right) topologies of the $\mathbf{N_2}$ dimer.

As can be observed in *Figure* 4-10, a more complex situation is found for **N** sequence. At least three signals above 15 ppm are observed, clearly indicating an equilibrium between different i-motif structures. Several dispersed signals are also found between 11 and 14 ppm. As only G:C base pairs (apart from hemiprotonated pairs) were expected for **N**, it may indicate that not all guanine residues are involved in "canonical" WC G:C base pairs and some residues are abnormally shielded.

Upon acidification, an alternative i-motif structures seems to be formed, based on the significant change observed in the NMR spectra at pH 5 (*Figure* 4-10, right). 2D-NMR studies are required for further analysis but, apparently, no WC G:C base pairs are formed and a new hemiprotonated imino signals pattern is observed.

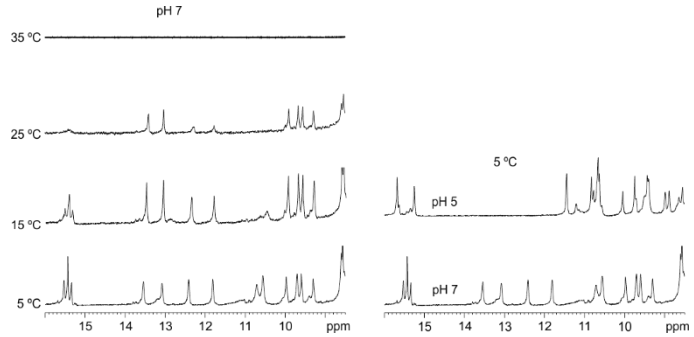


Figure 4-10. 1D ^{1}H-NMR melting series of **N** at pH 7 (left). Comparison of the 1D NMR spectra at pH 7 and pH 5 at 5 ºC (right). H_2O/D_2O 90:10, 10 mM phosphate buffer solution and 100 mM NaCl. [oligonucleotide] = 2.3 mM for pH 7 and 1.8 for pH 5.

4.2.2.2. **cN**, d<pCCGTTCCGTT>

In the case of **cN**, NMR analysis is highly facilitated by the symmetry provoked by chain cyclization. In this case, one unique dimeric mini i-motif structure containing G:C:G:C tetrads can be formed upon dimerization. Moreover, because of the double symmetry of the dimer, expected NMR signals for such structure are limited to five spin systems as C1≡C6, C2≡C7, G3≡G8, T4≡T9 and T5≡T10 (*Figure* 4-11).

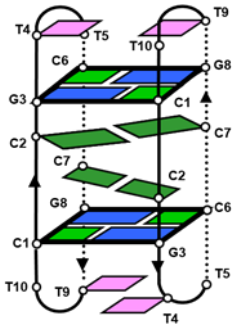


Figure 4-11. Expected dimeric structure for **cN**.

As can be observed in the 1D ^{1}H-NMR series at different pH, shown in *Figure* 4-12, formation of hemiprotonated C:C^{+} base pairs and WC G:C base pairs at neutral pH are supported by the signals observed above 15 and 13 ppm, respectively. As observed for **L** and **cL** sequences, the imino signal corresponding to G:C base pair formation disappears when lowering pH although is partially recovered when increasing temperature, even at acidic pH. Signals around 10 ppm probably can be attributed to amino protons of protonated cytosines.

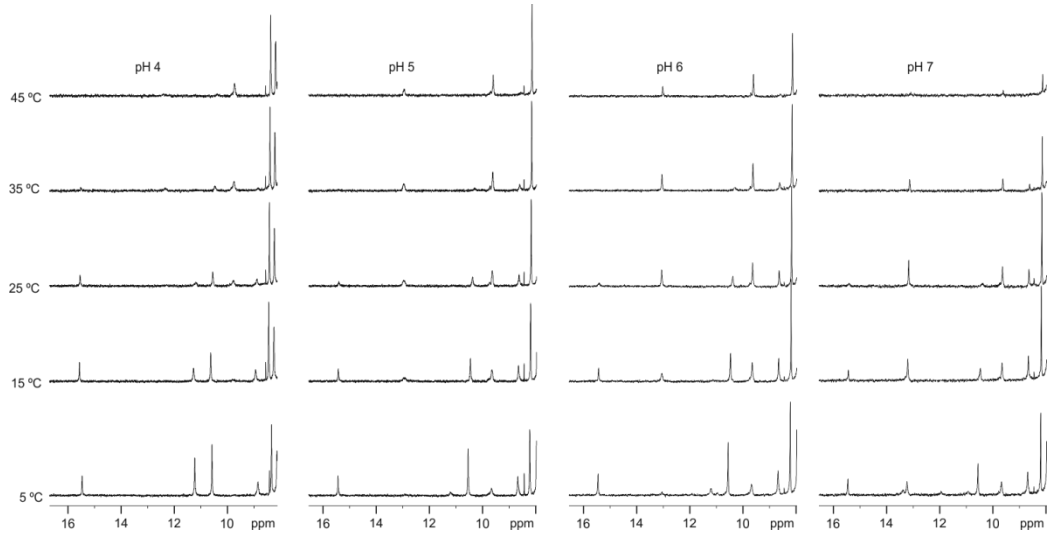


Figure 4-12. NMR melting series of **cN** at different pH values. H_2O/D_2O 90:10, 10 mM phosphate buffer solution and 100 mM NaCl. [oligonucleotide] = 0.8 M.

4.2.3. K-RELATED SEQUENCES

4.2.3.1. **K**, d(TCATTTCAT)

The attempt to observe the formation of slipped A:T:A:T tetrads in a mini i-motif context was approached by a double nucleotide substitution on the **M** sequence. Whereas the previously presented oligonucleotides were designed from the strategic replacement of the thymidine residues involved in the formation of the minor groove tetrads, the **K** sequence arises from substituting the guanine residues in positions three and eight by adenines. Analogously to **N**, for the $\mathbf{K_2}$ dimer formation two A:T:A:T tetrads would be expected independently of strand orientation within the structure (see *Figure* 4-13).

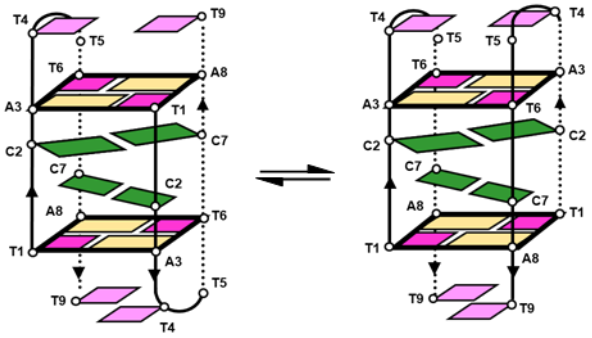


Figure 4-13. Expected head-to-tail (left) and head-to-head (right) topologies of the $\mathbf{K_2}$ dimer.

Remarkably, the NMR spectra recorded for **K** do not show i-motif formation under any experimental conditions. *Figure* 4-14 shows that, not even under strong acidic conditions (pH 4) that should theoretically favor the formation of i-motif structures, no characteristic signal corresponding to i-motif formation is observed. The only exchangeable proton signals are found between 11 and 12 ppm and their low intensity allow us to consider them negligible. Therefore, we must conclude that this oligonucleotide remains mainly disordered in solution.

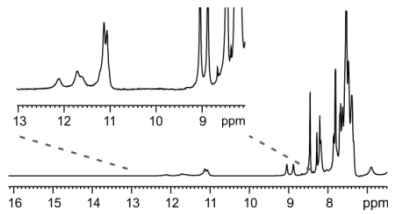


Figure 4-14. NMR spectrum of **K** at pH 4. H_2O/D_2O 90:10 10 mM phosphate buffer solution and 100 mM NaCl. [oligonucleotide] = 1.2 mM.

The conclusion obtained after these results is that A:T:A:T tetrads do not adopt the proper geometry to be compatible with the formation of mini i-motif structures. Considering the relative stability of A:T vs G:T base pairs,[222] we would not expect the nature of the capping base pair to be the reason behind the lack of structuration in **K** under any experimental conditions. Moreover, stacking interactions between for G or A with the hemiprotonated $C:C^+$ base pairs should be very similar. Thus, the reason why A:T:A:T tetrads are not formed in a mini i-motif context must be related to the fact that no slipped A:T:A:T minor groove tetrads can be formed in this case.

It is also worth to notice that without the favorable capping interactions provided by the minor groove tetrads at the ends of the $C:C^+$ stack, the stabilization of i-motif structures with no consecutive cytosine residues in their sequence seems to be unaffordable.

4.3. STABILITY

In order to evaluate the stability of the i-motifs formed by the different sequences, pH titrations and melting experiments have been carried out by using CD spectroscopy. Biophysical data corresponding to **M**, have been also included in this section for a proper comparison under the same experimental conditions.

It is important to highlight that, due to the dimeric nature of these structures, thermal stability strongly depends on oligonucleotide concentration. This fact prevents to directly compare the quantitative results

presented hereafter, and summarized in *Table* 4-2, with the results obtained from NMR experiments. However, a qualitative extrapolation of the relative stability provided by each tetrad should be feasible.

NAME	SEQUENCE	pH_T	T_m pH 6.0	T_m pH 5.0
M	d(TCGTTTCGT)	5.8	-	12.1
L	d(TCGTTCCGT)	6.3	13.8	22.9
N	d(CCGTTCCGT)	6.2	12.2	22.4

Table 4-2. Summary of the pH_T and T_m values calculated for the **M**, **L** and **N** sequences, Experimental conditions: 25 mM phosphate buffer, 100 mM NaCl and [oligonucleotide] = 20 μM

4.3.1. pH-TITRATION EXPERIMENTS

As it is described in *Section* 8.7.4 in Methods, the pH titration of the oligonucleotide samples monitored by CD spectroscopy allows to follow the folding/unfolding of their structures as a function of pH. Overlapped CD spectra recorded from acidic to basic pH (which translates into a denaturation of the sample) of **M**, **L** and **N** are shown in *Figure* 4-15. The spectra recorded at neutral pH are highlighted in red, showing how under these experimental conditions all sequences remain predominantly disordered displaying the typical CD profile of a random coil. At acidic pH values, however, all sequences show a high degree of structuration. Very interestingly, while **M** and **L** share a very similar CD profile with maximum bands between 260 and 270 nm and minimum bands around 240 nm, **N** exhibits a completely different profile with two maximum bands around 295 and 260 nm, similar to characteristic i-motif structure CD spectra. The CD profile found for the dimeric mini i-motifs of **M** and **cM** oligonucleotides at acidic pH is most probably due to a greater contribution of the minor groove tetrads than that of the $C:C^+$ stack. This information reveals that under these experimental conditions, while the i-motif structure of **L** resembles that of a mini i-motif structure, **N** seems to adopt a more classical-like i-motif structure. In order to elucidate both structures, a deep structural analysis will be presented in *Section* 4.4.

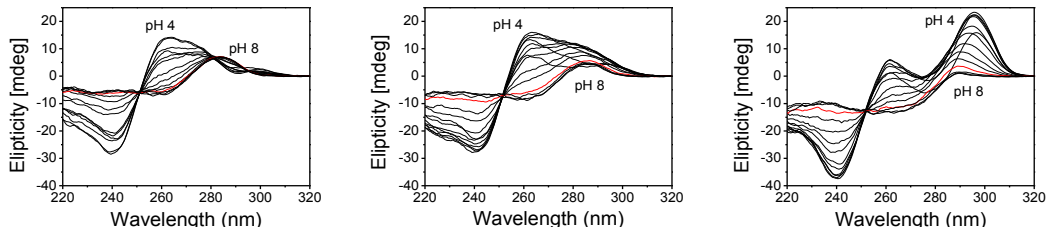


Figure 4-15. Superposition of CD spectra recorded at different pH values for the sequences **M** (left), **L** (center) and **N** (right) at 5 ºC and [oligonucleotide] = 20 μM. 25 mM phosphate buffer solution and 100 mM NaCl. Spectra at pH 7 are highlighted in red.

Moving on to quantitative results, from the spectra showed in *Figure* 4-15, sigmoidal curves can be obtained by plotting ellipticity values from a fixed wavelength as a function of pH (*Figure* 4-16). The chosen wavelength values are those of maximum ellipticity (265 nm for **M** and **L** and 295 nm for **N**). From acidic to basic pH values, the obtained curves display an initial strong increase of ellipticity that corresponds to the deprotonation of the pH-sensitive nucleobases. Then, a plateau of maximum ellipticity indicates the formation of stable secondary structures that ultimately drops to minimum values of ellipticity indicating the denaturation of the structure.

CD vs pH data can be fitted to a sigmoidal function by using the Boltzmann equation (*Section* 8.7.5), in order to obtain transitional pH (pH_T) values. As it was predictable, all pH_T values are lower than 7, observing, however, a significant increase of those of **L** and **N** compared to that of **M** (6.3, 6.2 and 5.8, respectively) (see *Table* 4-2). This information strongly supports that the substitutions performed on the original sequence of **M** render i-motif structures that are stable over a wider range of pH.

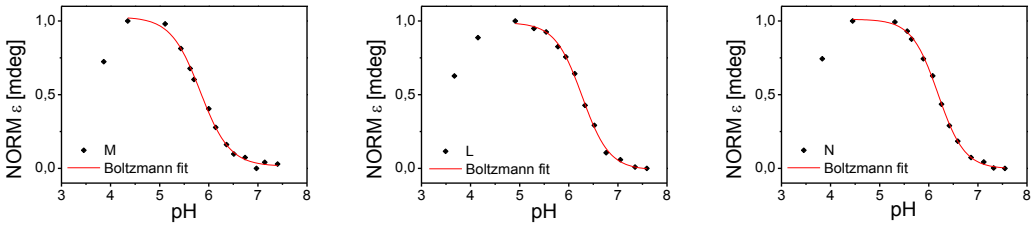


Figure 4-16. CD-monitored pH titration curves of **M** (left), **L** (center) and **N** (right) at 5 ºC. 25 mM phosphate buffer, 100 mM NaCl and [oligonucleotide] = 20 µM. Boltzmann fit represented in red for the denaturalization of the structures.

4.3.2. THERMAL STABILITY

The very low stability observed in the previous section for **M**, **L** and **N** sequences at neutral pH and 20 µM oligonucleotide concentration prompted us to evaluate their thermal stability under acidic conditions. However, CD-monitored melting curves and CD spectra at different temperatures provided still very low stability values. *Figures* 4-17 to 20 show CD spectra at different temperatures and CD-monitored melting curves of **M**, **L** and **N** at pH 6 and 5. The T_m values obtained from melting curves were calculated using a Boltzmann fit and are summarized in *Table* 4-2.

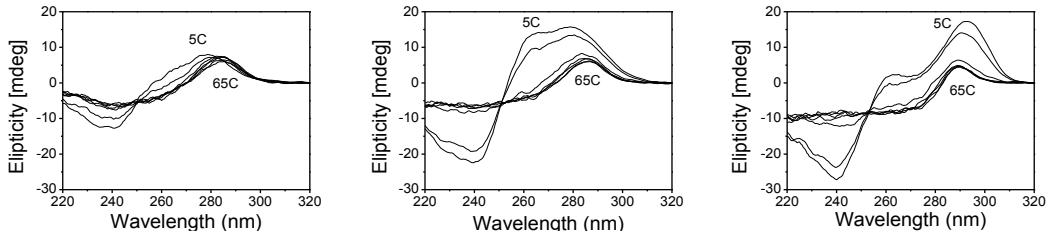


Figure 4-17. Superposition of CD spectra recorded at different temperatures (5, 10, 25, 35, 45, 60 and 65 ºC) for the sequences **M** (left), **L** (center) and **N** (right) at pH 6 and [oligonucleotide] = 20 µM. 25 mM phosphate buffer solution and 100 mM NaCl.

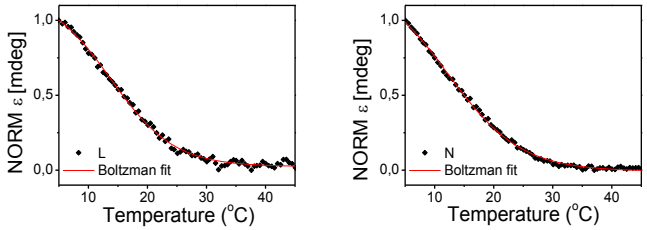


Figure 4-18. CD-monitored melting curves of the sequences **L** (left) and **N** (right) at pH 6 and [oligonucleotide] = 20 µM. 25 mM phosphate buffer solution and 100 mM NaCl.

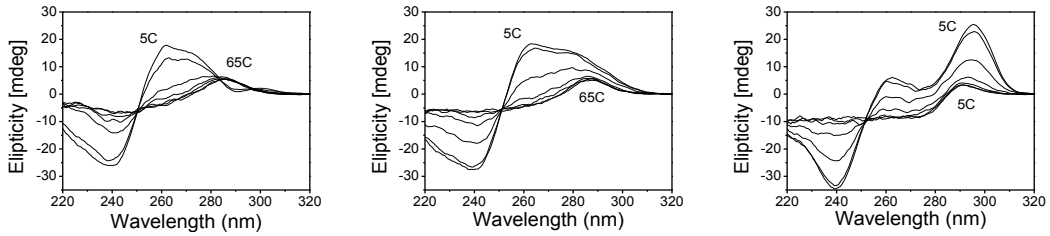


Figure 4-19. Superposition of CD spectra recorded at different temperatures (5, 10, 25, 35, 45, 60 and 65 ºC) for the sequences **M** (left), **L** (center) and **N** (right) at pH 5 and [oligonucleotide] = 20 µM. 25 mM phosphate buffer solution and 100 mM NaCl.

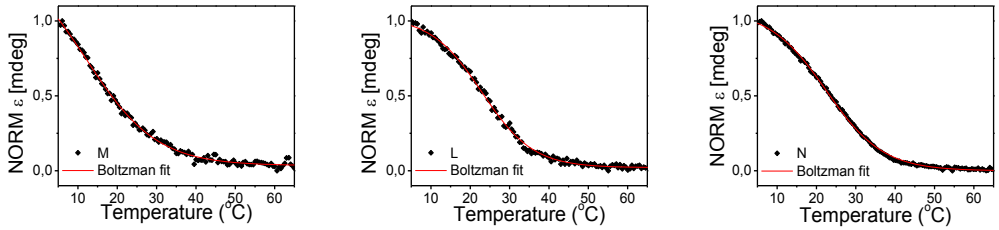


Figure 4-20. CD-monitored melting curves of the sequences **M** (left), **L** (center) and **N** (right) at pH 5 and [oligonucleotide] = 20 μM. 25 mM phosphate buffer solution and 100 mM NaCl.

Overall, the significant information extracted from these melting experiments is rather poor. At pH 6, concordantly to its pH_T value, the **M** sequence is found unstructured and thus does not exhibit a melting profile while **L** and **N** show very low T_m values. As expected for i-motif structures, the stability of the motifs increases significantly (+ 10 ºC aprox.) from pH 6 to pH 5. However, even at pH 5, the obtained T_m values reflect that these structures would be only partially folded at room temperature. Finally, and most significantly, the thermal stability data confirm the trend observed in the previous section, suggesting that **L** and **N** i-motif structures are more stable than that of the original sequence **M**.

4.3.3. STABILITY OF CYCLIC ANALOGUES

The stability data of **cL** and **cN** presented in *Table* 4-3 was recorded by M. Garavís from Prof. C. González's group in Madrid by using CD spectroscopy and the same methodology described in the experimental section of this thesis.

NAME	SEQUENCE	pH_T	T_m @pH6.0	T_m @pH5.0
cL	d<pTCGTTCCGTT>	6.1	34.6	45.1
cN	d<pCCGTTCCGTT>	6.1	41.3	52.6

Table 4-3. Summary of the pH_T and T_m values calculated for the **cL** and **cN** sequences, Experimental conditions: 25 mM phosphate buffer, 100 mM NaCl and [oligonucleotide] = 50 μM

In *Figure* 4-21, the dependence of the cyclic structures with the pH is presented. Most surprisingly, the CD profile of **cN** (which is basically the same than that of **cL**) does not match the one obtained for its linear analogue, but resembles the ones obtained for **M** or **L**. This data points to a different pH behavior of **N** and **cN**. Further characterization from 2D-NMR studies will be crucial to elucidate the nature of these structures.

As shown in *Table* 4-3, the pH_T values for **cN** and **cL**, obtained from the sigmoidal fit of the maximum ellipticity values plotted against pH values, are the same, and are comparable to those obtained for their linear analogues.

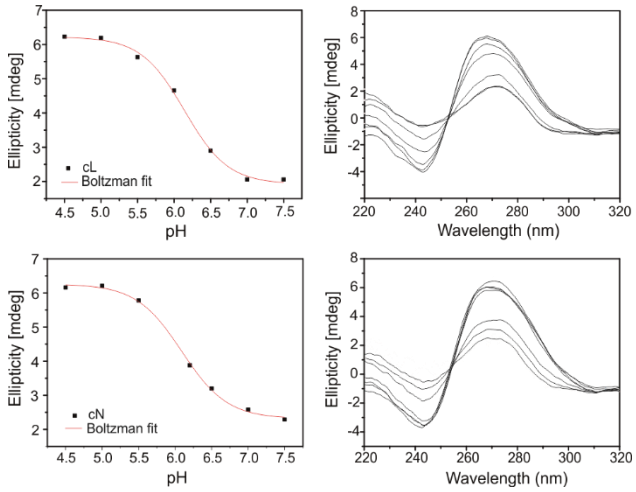


Figure 4-21. Superposed CD spectra recorded at different pH values (right) and sigmoidal curves of CD (265 nm) vs pH (left) of the sequences **cL** (top) and **cN** (bottom). [oligonucleotide] = 50 μM. 25 mM phosphate buffer solution and 100 mM NaCl.

Remarkably, the melting experiments recorded for the cyclic sequences afford much higher T_m values than those obtained for the original linear sequences (see *Figure* 4-20). However, some considerations must be taken into account before comparing these results. First, the intrinsic nature of the cyclic oligonucleotides provides improved thermal stability to their secondary structures. Second, the experimental difference in the oligonucleotide concentration of the samples (20 μM for **L** and **N** and 50 μM for **cL** and **cN**) prevents a direct comparison of the results due to the dimeric nature of the structures. Finally, the fact that for **cL** a coexistence of alternative species at acidic pH might be able to explain the notable difference in stability compared to **cN**.

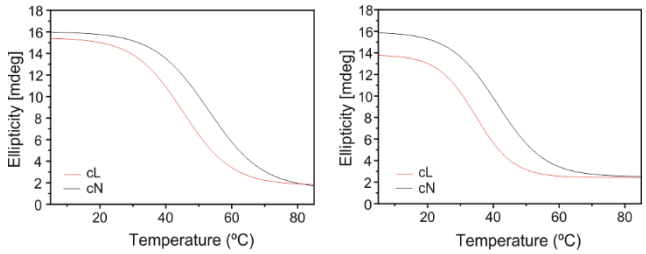


Figure 4-22. CD-monitored melting curves of the sequences **cL** and **cN** at pH 5 (left) and pH 6 (right) [oligonucleotide] = 50 μM. 25 mM phosphate buffer solution and 100 mM NaCl.

4.4. STRUCTURAL CHARACTERIZATION

To gain more insight into the folding of these sequences and explain the reasons for the formation and the stability data reported in the previous sections, we undertook a more detailed structural study using two-dimensional NMR spectroscopy.

4.4.1. L-RELATED SEQUENCES

4.4.1.1. THE SEQUENCE **L** FOLDS INTO A MAJOR HEAD-TO-TAIL DIMER AT pH 7

The recorded 2D NMR spectra at pH 7 for the sequence **L** exhibit signals corresponding to a major species (10 spin systems) and some other signals corresponding to unstructured or minor species. Focused on the major species, sequential H2'/H2"-H6/H8 cross-peaks in the non-exchangable proton region of the spectra allow to unequivocally follow two connected fragments: T1-C2-G3-T4 and C6-C7-G8-T9 (see *Figure* 4-23). T5 residue does not show any inter-residual cross-peaks. According to imino-amino cross-peaks, C2 and C7 form hemiprotonated C:C^+, as both couple of amino protons show cross-peak with the protonated imino signal found at 15.5 ppm (see *Figure* 4-24). Interestingly, the amino protons of C7 appear significantly more downfield shifted (10.45 and 8.04 ppm) than those of C2 (8.70 and 7.37 ppm), indicating that they are exposed to a different chemical environment. In principle, this experimental data would be compatible with two scenarios: the formation of two hemiprotonated C2:$C7^+$ base pairs in a head-to-tail folded structure (*Figure* 4-5, left), or the formation of symmetric C2:$C2^+$ and C7:$C7^+$ base pairs in a head-to-head structure (*Figure* 4-5, right). In both cases, amino protons of C2 and C7 may be differently shielded as C2 and C7 would be stacked onto different types of base pairs (WC G:C or G:T mismatched base pairs). In the case of the head-to-head structure, however, differences for the protonated imino signal of each pair would be expected. Therefore, the chemical shifts observed point to a head-to-tail orientation.

The C6 residue was identified to form a WC G:C base pair with G8 residue based on amino contacts with the imino signal of G8 at 13.75 ppm. G8 stacking interactions with C7 (H1G8-H42C7, H42C7-H21G8), unambiguously support the formation of C6:G8 W-C base pair. Although the expected cross-peak H3T-H1G is not observed at pH 7, G3 exchangeable protons exhibit chemical shifts compatible to G:T base pair formation: H1G3 at 10.77 ppm and H22/H21G3 at 8.63/7.63. Furthermore, G3 also show stacking cross-peaks with C2 (H1G3-H5/H41/H42C2). T1 imino proton signal at 11.59 ppm exhibits a broad profile probably related to the residue being in a terminal position. With the G:C and G:T base pairs characterized, the formation of slipped G:C:G:T tetrads is determined after confirming that G3 and G8 residues interact through their Hoogsteen side, since imino H1 protons of G3 (10.76 ppm) and G8 (13.73 ppm) show cross-peaks with H1' protons of G8 and G3, respectively. The head-to-tail orientation is also supported by amino-amino cross-peaks between G3 and G8 residues and H42C7-H2'/H2"C6 and H41C7-H1'C6, which occur between intercalated base pairs through the major groove of the i-motif.

Additional NOEs, such as MeT4-H8/H1'G3 and MeT9-H8/H1'G8, indicate stacking interactions between G3-T4 and G8-T9.

Assigned protons of L at pH 7 are listed in *Table App* 1.

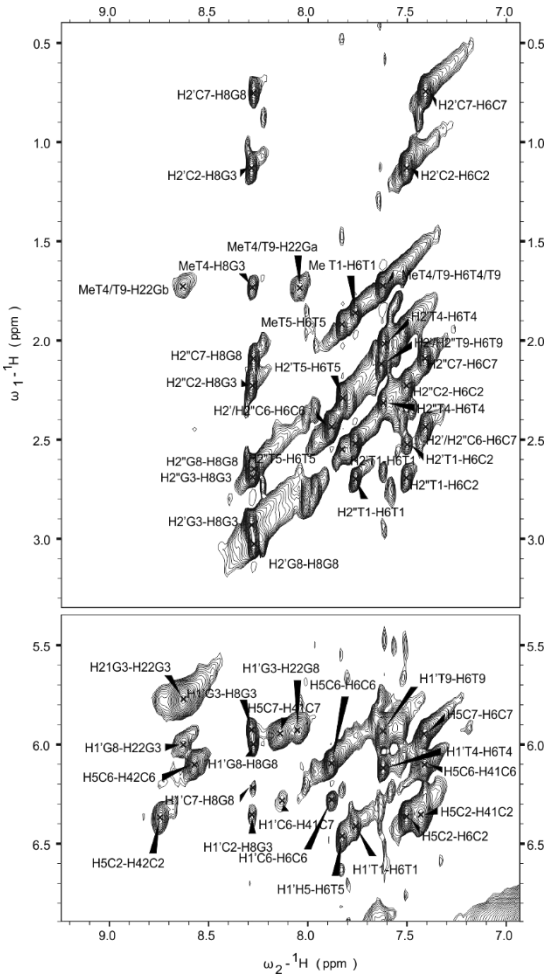


Figure 4-23. Non-exchangeable protons region of the NOESY spectra (150 ms) of **L** at pH 7 and 5 ºC. H_2O/D_2O 90:10, 10 mM phosphate buffer, [oligonucleotide] = 1.5 mM.

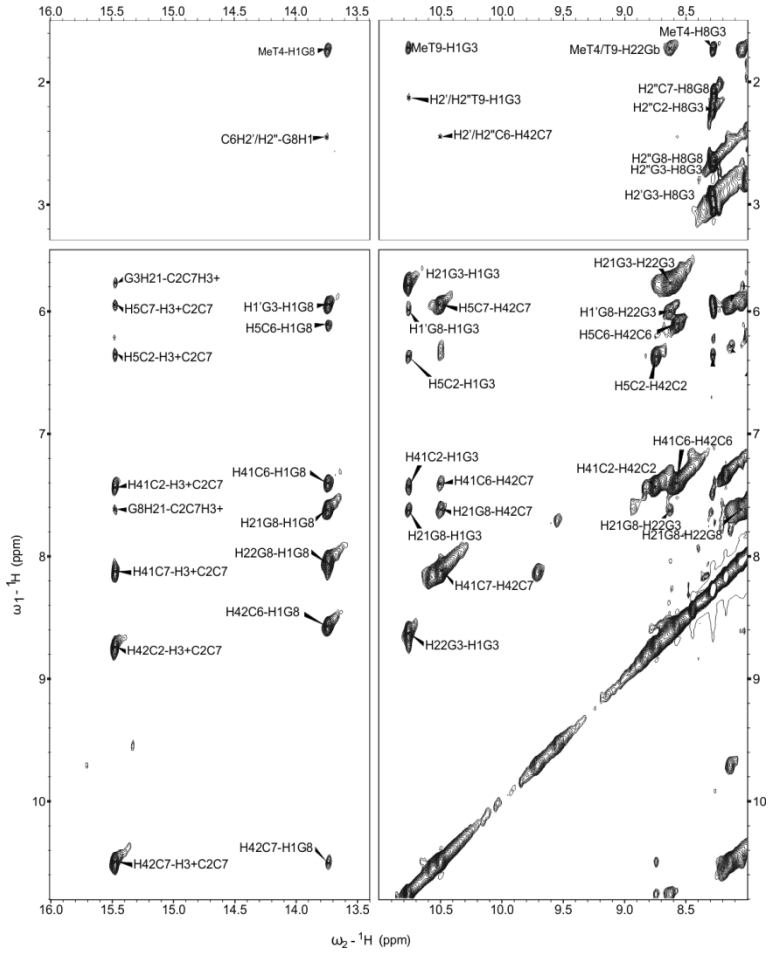


Figure 4-24.- Exchangeable protons region of the NOESY spectra (150 ms) of **L** at pH 7 and 5 ºC. H_2O/D_2O 90:10, 10 mM phosphate buffer, [oligonucleotide] = 1.5 mM.

4.4.1.2. THE HEAD-TO-HEAD MOTIF APPEARS AT ACIDIC pH VALUES AS A MINOR SPECIES

Most of the spectral features observed at pH 7 remain unaltered at pH 6. The most notable difference is the increased intensity of the signal at 11.59 ppm corresponding to a thymine imino proton. Unfortunately this signal does not show any NOE cross-peaks that could assess the formation of a G:T base pair. At pH 6, two small additional signals at 15.72 and 15.31 ppm are observed, indicating the formation of additional i-motif species. Both imino protons show crosspeaks with only a pair of amino protons, indicating the formation of hemiprotonated $C:C^+$ base pairs between equivalent cytosines ($C2:C2^+$ and $C7:C7^+$). These signals most likely correspond to a minor population of a dimeric head-to-head i-motif structure. However, the absence of further well-defined cross-peaks to connect these extra $H3^+$ signals impedes the full characterization of this minor species.

The 2D NMR spectra recorded at pH 5 shown in *Figure* 4-26, maintain the pattern of cross-peaks described by the previously discussed spectra with an increasing presence of the dimeric head-to-head i-motif structure. Nevertheless, the head-to-tail structure seems to remain as the major species in all cases. At pH 5, the imino signal at 11.59 ppm could be unambiguously assigned to T1 and remarkably shows a NOE H3T1-H1G3 cross-peak confirming the formation of a G:T base pair (see *Figure* 4-26). Unfortunately, under these experimental conditions, the imino signal of G8 is only barely observed.

More structural details can be obtained from NMR data consistent with the head-to-head dimer at pH 5:

- Two additional signals in the 15 ppm region that show cross-peaks with only one amino protons pair, indicating C:C$^+$ formation between equivalent cytosines.
- Sequential H8-H2'/H2" cross-peaks for two C-G steps are clearly observed for the minor species: C2-G3 and C7-G8.
- One H3T-H1G cross-peak is observed, indicative of G:T base pair formation. Guanine residue implied in the G:T exhibits stacking cross-peaks with one of the hemiprotonated cytosines.
- H3T signal shows cross-peaks with H2'/H2" protons of C6 These cross-peaks correspond to contacts between pyrimidine residues in the tetrads across the major groove. In the case of the head-to-head structure this contacts are between thymines (see *Figure* 4-26).
- G:C base pair formation is not observed but C6 seems to participate in hydrogen bonding as H5-H42C6 cross-peak is found. A signal at 13.14 ppm it is observed in the diagonal that might correspond to an imino proton of G:C W-C, but it does not show any cross-peaks.

Chemical shifts corresponding to head-to-tail and head-to-head species at pH 5 are given in *Tables App* 2 and 3, respectively. In the case of the head-to-head species only few residues could be unambiguously assigned. Schemes of the proposed head-to-tail and head-to-head structures are given in *Figure* 4-25.

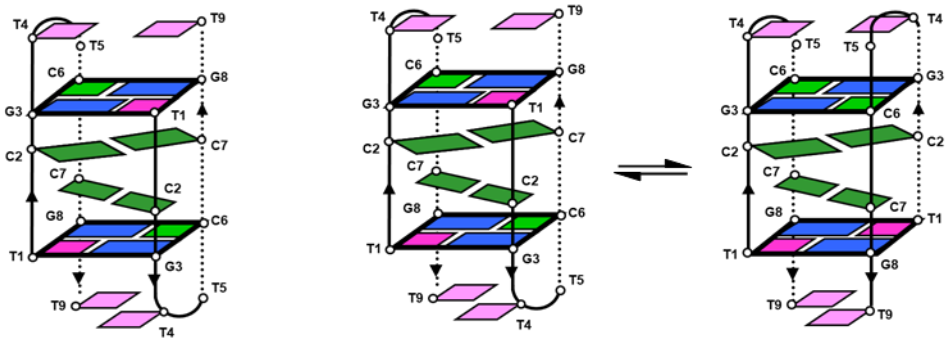


Figure 4-25.- Model of the proposed main structures found at neutral (left) and acidic pH (right).

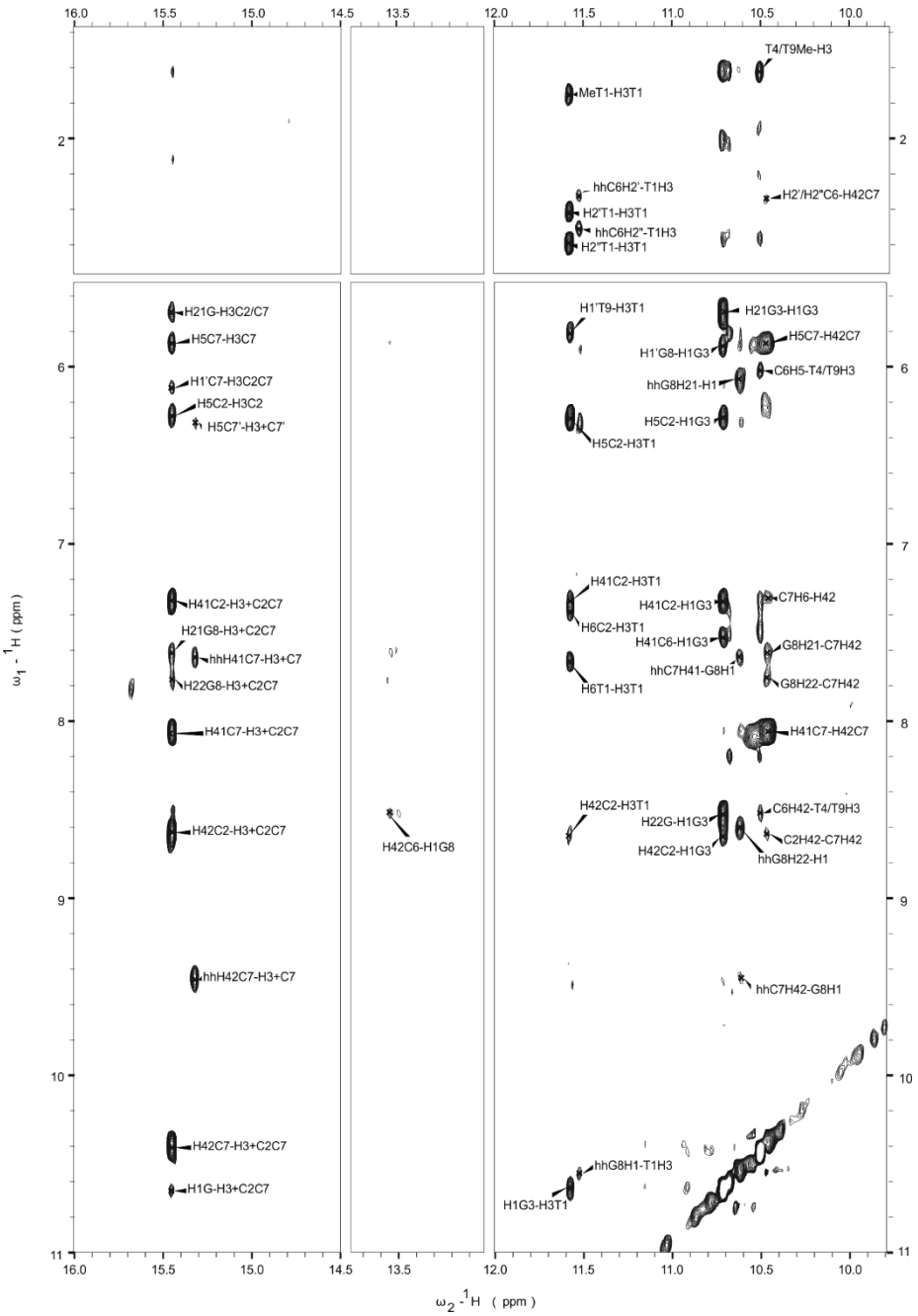


Figure 4-26.- Exchangeable protons region of NMR spectra (150 ms) of **L** at pH 5 and 5 ºC. H_2O/D_2O 90:10, 10 mM phosphate buffer, [oligonucleotide] = 1.5 mM. Signals for the head-to-head species are designated as “hh”.

The reasons for the formation of head-to-head species at acidic pH are not clear. It may be related to the pH-dependence of tetrads containing WC G:C base pairs and the role that protonated cytosines involved in these base pairs can play. The fact that, even at acidic pH, the major species correspond to the head-to-tail one supports that mixed G:C:G:T tetrads are resistant to acidic pH (i.e. cytosine protonation). Protonation of C6 in this case cannot afford an extra $C:C^+$ base pair and, when protonated, a destabilization of the tetrad may occur. As no signal of C6 protonation is observed in the NMR spectra, it can be possible that structural effects are lowering the pK_a value expected for this cytosine. However, if partial protonation of C6 occurs, this might favor the structure in which, at least one G:T:G:T non pH-dependent minor groove tetrad is formed.

4.4.1.3. THE CYCLIC ANALOGUE **cL** ALSO PREFERS THE FORMATION OF G:C:G:C MGT

Nice 2D NMR spectra of **cL** are obtained at pH 6 as shown in *Figure* 4-27 and *Figure* 4-28. Under these experimental conditions the major species (10 spin systems) is also the i-motif structure. Besides the signals corresponding to the formation of hemiprotonated $C:C^+$ base pairs, signals corresponding to the formation of G:T and G:C base pairs are clearly observed. The formation of G:C:G:T minor groove tetrads is confirmed by the presence of the H1T-H3G and H1/H21/H22G-H1'G cross-peaks. As for the linear sequence, one imino signal that show cross-peaks with two pairs of amino protons is observed at 15.54 ppm, indicating the formation of hemiprotonated $C:C^+$ base pairs between non-equivalent cytosines ($C2:C7^+$). These $C2:C7^+$ base pairs are only consistent with an i-motif structure capped by G:C:G:T tetrads. Consequently, the two macrocycles are antiparallelly arranged and mutually rotated 180º (*Figure* 4-7, right). According to this orientation, C6 is forming a WC G:C base pair with G8, and G3 is forming a G:T base pair with T1. The complete assignment could be accomplished on the basis of sequential and stacking cross-peaks (H2'/H2"/H1'C2-H8G3, H2'/H2"/H1'C7-H8G8, MeT4-H1'G3, MeT9-H1'G8, MeT1-H1'T10). The remaining spin system was assigned to T10. Chemical shifts are given in *Table App* 4

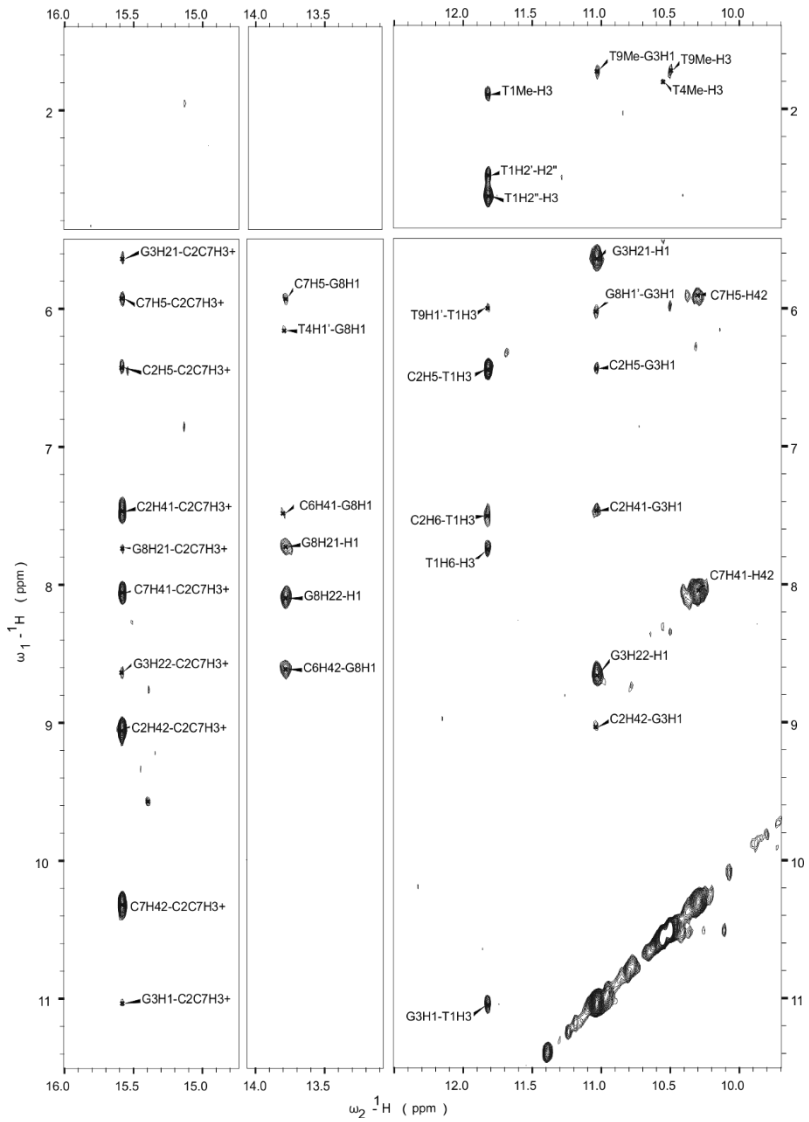


Figure 4-27.- Exchangeable protons region of the NOESY spectra (150 ms) of **cL** at pH 6 and 5 ºC. H_2O/D_2O 90:10, 10 mM phosphate buffer, [oligonucleotide] = 0.6 mM.

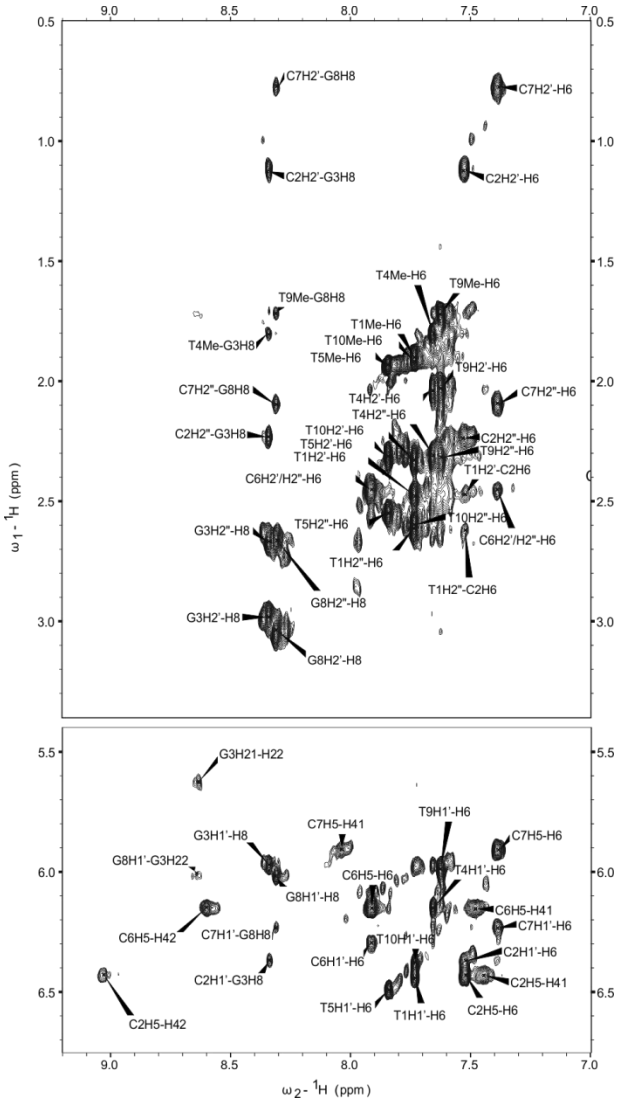


Figure 4-28. Non-exchangeable protons region of the NOESY spectra (200 ms) of **cL** at pH 6 and 5 ºC. H_2O/D_2O 90:10, 10 mM phosphate buffer, [oligonucleotide] = 0.6 mM.

Upon lowering pH, some changes are observed, but the main species appears to be the same i-motif structure with the same strand orientation as that found at pH 6. The cross-peaks pattern corresponding to the hemiprotonated $C:C^+$ base pairs indicates the formation of $C2:C7^+$ pairs, and the formation of G3:T1 base pairs is clearly observed. Amino protons of G3 show cross-peaks with H1' proton of G8, indicating G:G interaction through the minor groove. Although, imino and amino protons of G8 are not observed and the formation of WC G8:C6 base pair could not be confirmed, the chemical shifts of the amino protons of C6 are those of a WC paired cytosine (8.64 and 7.68 ppm).

As for the linear, small signals corresponding to a minor species appeared at pH 5. A small signal at 15.40 ppm that shows cross-peaks with one pair of amino protons indicates that this minor species probably corresponds to a dimeric i-motif structure where equivalent residues are confronted (see *Figure* 28).

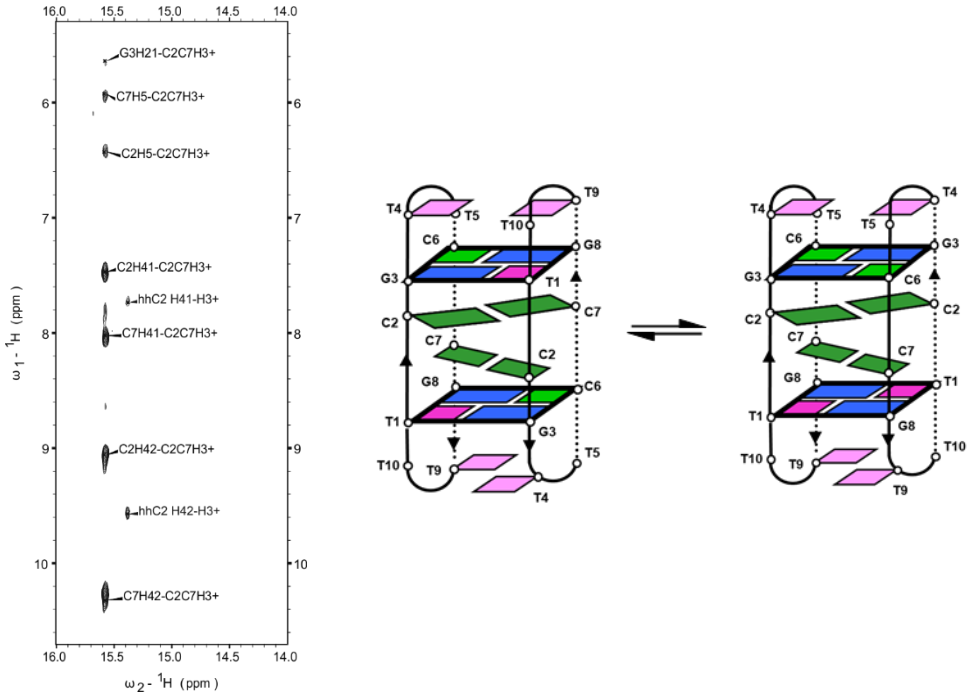


Figure 28. Exchangeable protons region of the NOESY spectra (150 ms) of **cL** at pH 5 and 5 ºC. H_2O/D_2O 90:10, 10 mM phosphate buffer, [oligonucleotide] = 0.6 mM (left). Minor species has been designated as "hh". Schematic representation of the orientations of the major and the minor species characterized for **cL** (right).

4.4.2. N-RELATED SEQUENCES

4.4.2.1. A 1:1 MIXTURE OF TOPOLOGIES WITH G:C:G:C MGT IS OBSERVED FOR **N** AT pH 7

NMR spectra of **N** at pH 7 show a complex equilibrium of at least four species similarly populated. In the TOCSY spectrum, up to eleven H5-H6 cross-peaks are observed and, some of them, clearly correspond to more than one overlapped cytosine. However, good signal dispersion is found for some of the non-exchangeable protons and, importantly, in the exchangeable protons region. All the exchangeable proton signals seem to correspond to two structured i-motif species.

Three signals are found in the $C:C^+$ base pairs region as it is observed in the 1D NMR spectra (*Figure* 4-10). Two of the imino $C:C^+$ signals (15.53 and 15.33 ppm) correspond to two homo-pairs whereas the third signal (15.43 ppm) corresponds to a $C:C^+$ base pair between non-equivalent residues. Starting from these signals two of the species can be identified as two dimeric mini i-motifs with the two possible

strand orientation: head-to-tail and head-to-head. Aromatic protons of residues corresponding to these two species, exhibit good dispersion and could be assigned on the basis of sequential Ar-H1'/H2'/H2" cross-peaks. In both cases it was possible to follow C1-C2-G3-T4 and C6-C7-G8T9 fragments. As found for dimeric mini i-motif of M, the two species seem to have similar stability as they exhibit an approximately 1:1 ratio.

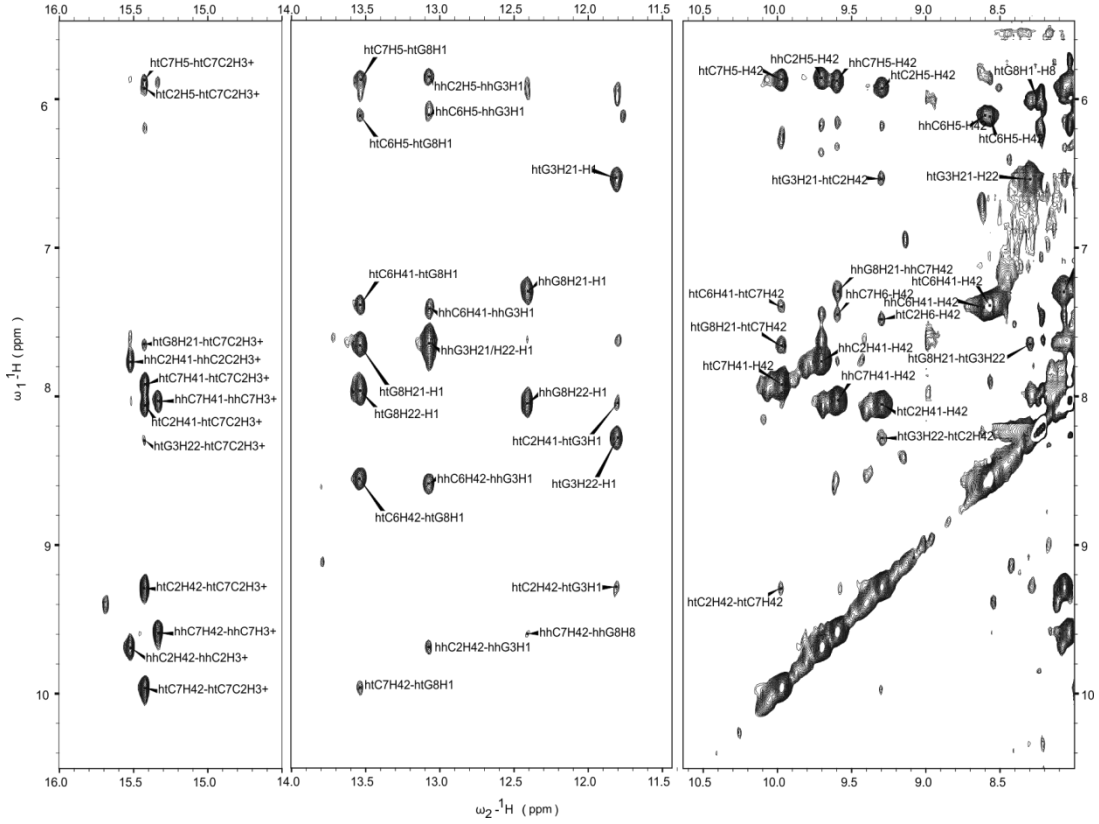


Figure 4-29. Exchangeable protons region of the NOESY spectra (150 ms) of **N** at pH 7 and 5 ºC. H_2O/D_2O 90:10, 10 mM phosphate buffer, [oligonucleotide] = . hh labels correspond to the signals of the head-to-head dimer. ht labels correspond to the signals of the head-to-tail dimer.

In both cases, formation of WC G:C base pairs are also observed. Four guanine imino signals are found in the region of bonded guanines but, as clearly observed in *Figure* 4-10, dispersed in a wide chemical shift range. For each species, one guanine imino proton is found in the typical region of G:C base pairs (13.55 and 13.08 ppm) and exhibits expected H1-H41/H42 cross-peaks with bonded cytosines. The other two signals (one for each species) are found upfield (12.41 and 11.81). These signals exhibit cross-peaks with own amino protons (downshifted) but not with presumably bonded cytosines. Interestingly, these cytosines also exhibit downshifted amino protons signals and amino-H1' contacts between guanine residues are found. All this data would be compatible with the formation of a G:C:G:C

tetrad in which one WC G:C base pair is formed and the other base-pair is not totally formed o it is formed a “G:T”-like base pair. All this data supports the formation of a head-to-tail and a head-to-head mini i-motifs capped by a sort of slipped G:C:G:C minor grove tetrads.

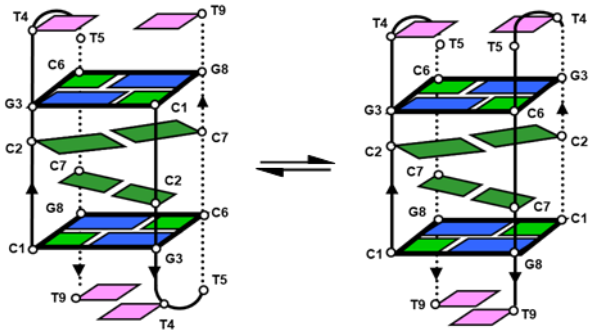


Figure 4-30. Head-to-tail and head-to-head equilibrium found for **N** at pH 7.

In addition to these two species, more non-exchangeable signals are found. One of the extra H5-H6 cross-peak seems to integrate four cytosines. Two of these cytosines exhibit clear sequential Ar-H1’/H2’/H2” cross-peaks, indicating that it is also structured, although no exchangeable proton signal is found for this species. No sequential cross-peaks are found for the two other H5-H6 cross-peaks, that would correspond to two cytosines each.

Chemical shifts of head-to-tail and head-to-head species are given in *Tables App* 5 and 6, respectively.

4.4.2.2. A HEAD-TO-TAIL DIMER WITH EXTRA C:C$^+$ BASE PAIRS IS FORMED AT ACIDIC pH

When lowering pH, up to eight cytosine H5-H6 cross-peaks of different intensities are observed, evincing an equilibrium between two species. According to H6 chemical shift, these cytosines are allocated in two groups, four of them exhibit H6 chemical shift of around 8 ppm, whereas the other four residues are found upshifted at around 7.5 ppm. Regarding the major species, upshifted cytosines show sequential cross-peaks with guanine residues, indicating that these signals belong to C2/C7 residues, whereas the other two signals may correspond to C1/C6 residues. Four protonated imino signals are observed above 15 ppm (15.73, 15.70, 15.65 and 15.26) (see *Figure* 4-31). Protonated imino signal at 15.26 ppm exhibits cross-peaks with cytosine residues C2/C7, whereas the signal at 15.65 ppm, the most intense, corresponds to an imino signal that exhibit cross-peaks with amino protons of residues C1/C6. These two signals have been assigned to the major species in which formation of four hemiprotonated C:C$^+$ base pairs was expected. Interestingly intra-residual amino-H2’/H2” contacts have been observed for

C1 and C6 residues, indicating that these residues are forming C1:C6^{+} base pairs located in the centre of the stack and facing their 3'-sides (see *Figure* 4-32).

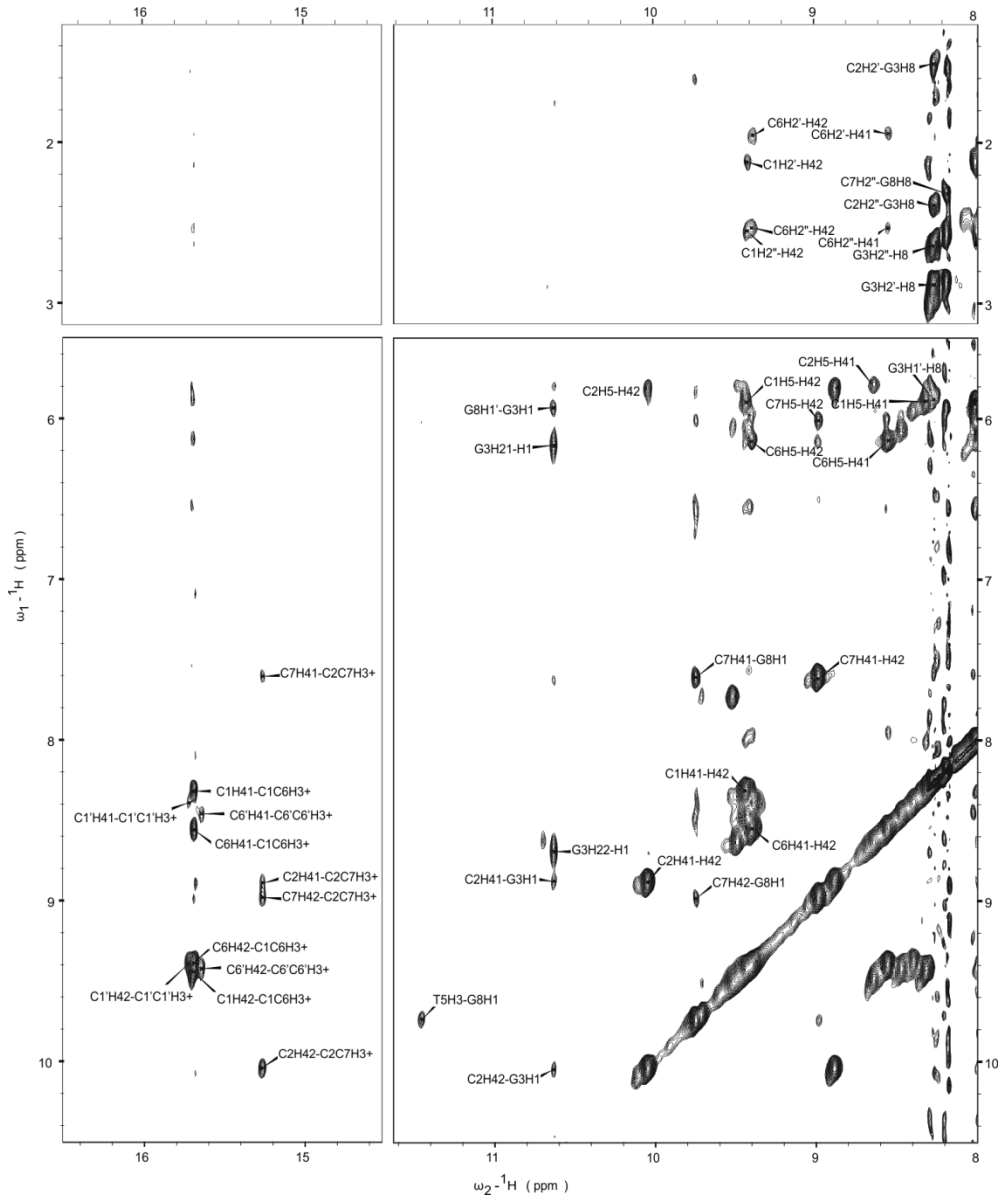


Figure 4-31. Exchangeable protons region of the NOESY spectra (150 ms) of **N** at pH 5 and 5 ºC. H_2O/D_2O 90:10, 10 mM phosphate buffer, [oligonucleotide] = 1.5 mM.

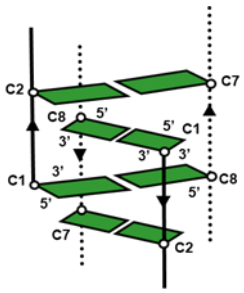


Figure 4-32. Schematic representation of the stacking order of the C:C$^+$ base pairs of the structure of **N** at pH 5.

A non-expected thymine imino signal at 11.46 ppm was also observed, corresponding to a G:T base pair that could be assigned to a T5:G8, due to stacking interactions with C7. H1 proton of G8 also showed a cross-peak with H1' proton of G3, indicating that, although the formation of a tetrad is not feasible in this case, there do exist minor groove interaction between both guanine residues.

Altogether, these data indicate that, at acidic pH, protonation of cytosine residues involved in WC G:C base pairs at neutral pH occurs, resulting in an alternative i-motif structure where four C:C+ base pairs are formed. According to signal intensities, imino proton of C2:C7$^+$ base pairs are not so solvent protected than that of C1 and C6, consistent with the fact that no complete tetrad can be formed at the ends of the stacks forming, however, some sort of G:G:T triad. In *Figure* 4-333 a proposal for the **N_2** dimer structure at acidic pH is shown.

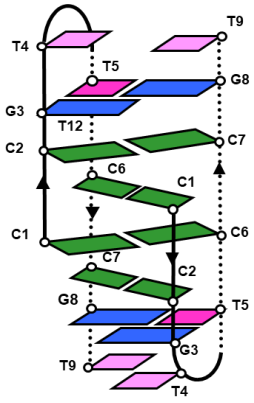


Figure 4-33. Schematic representation of the proposed head-to-tail dimer formed by **N** upon acidification.

Chemical shifts of the proposed species at acidic pH are given in *Table App* 7.

4.4.2.3. THE CYCLIC NATURE OF **cN** PREVENTS THE FORMATION OF ALTERNATIVE SPECIES

NMR spectra of d<pCCGTTCCGTT> clearly indicates the formation of an i-motif at pH 7, as an imino signal corresponding to the formation of hemiprotonated C:C$^+$ base pair is observed at 15.45 ppm. Some additional signals corresponding to unstructured species are also observed in the non-exchangeable region, probably due to the low oligonucleotide concentration of the sample. This indicates that the equilibrium between folded and unfolded species is slow in the NMR time scale (as usual in cyclic oligonucleotides).

For the major species only five spin systems are observed, indicating the formation of a symmetric dimer. In this case, as previously mentioned, the two possible strand orientations are equivalent. Two different types of cytosine residues are observed, one involved in the formation of C:C$^+$ base pairs (C2/C7), whose amino protons show cross-peaks with the imino signal at 15.45 ppm, and one forming WC G:C base pairs (C1/C6). One imino signal of guanine is observed at 13.23 ppm and shows cross-peaks with amino/H5 protons of C1/C6. Amino protons of guanine show cross-peaks with H1', indicating the interaction of the G:C base pairs through the minor groove side. Two different types of thymine residues are observed, one stacked above guanine residues (assigned to T4/T9) and other completely disordered (assigned to the only remaining thymine: T5/T10). All the NMR data is consistent with the formation of an i-motif structure with two hemiprotonated base pairs involving residues C2 and C7 and the formation of two capping G:C:G:C slipped MGT involving C1/C6 and G3/G8 residues. T4 and T9 are stacked onto the MGT, whereas T5 and T10 residues are exposed to the solvent.

At lower pH, the signals corresponding to unstructured species disappear and only five spin systems are observed. As for the other two related sequences, the NMR spectra are very similar as those obtained at pH 7, indicating the formation of the same structure. At pH 5 two different cytosines are observed: one forming hemiprotonated base pairs; and the other with amino signals at low field (at least one observed at 8.73 ppm). Imino and amino signals of guanine residues are not observed, but cross-peaks indicating a perfect stack C1-C2-G3-T4 are observed in the non-exchangeable protons region. Two different types of thymine residues are also observed, one stacked onto guanine residues and the other exposed exhibiting only intra-residual cross-peaks and probably to the solvent.

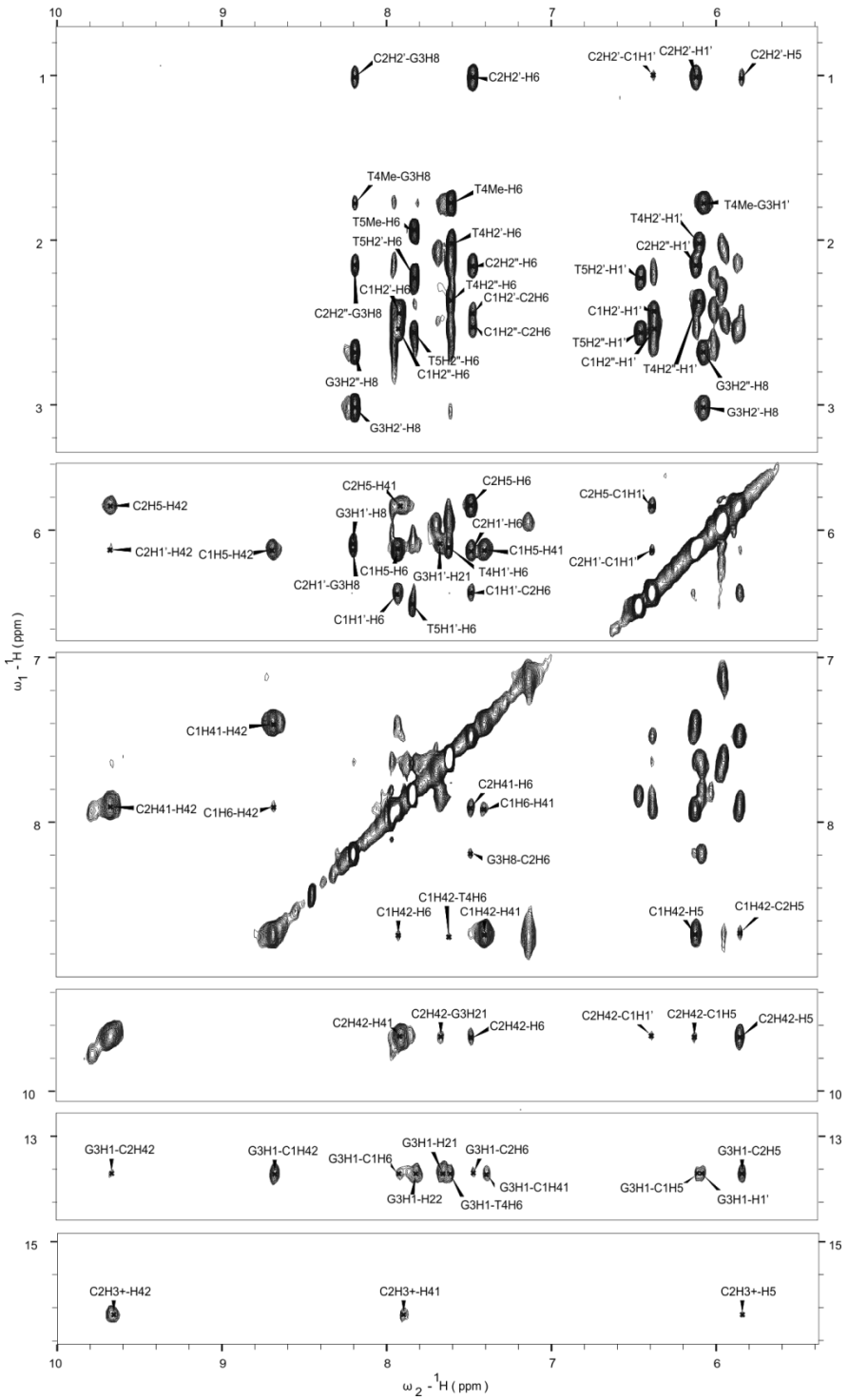


Figure 4-34.- Exchangeable and non-exchangeable protons region of the NOESY spectra (150 ms) of **cN** at pH 7 and 5 ºC. H_2O/D_2O 90:10, 10 mM phosphate buffer, [oligonucleotide] = 0.8 mM.

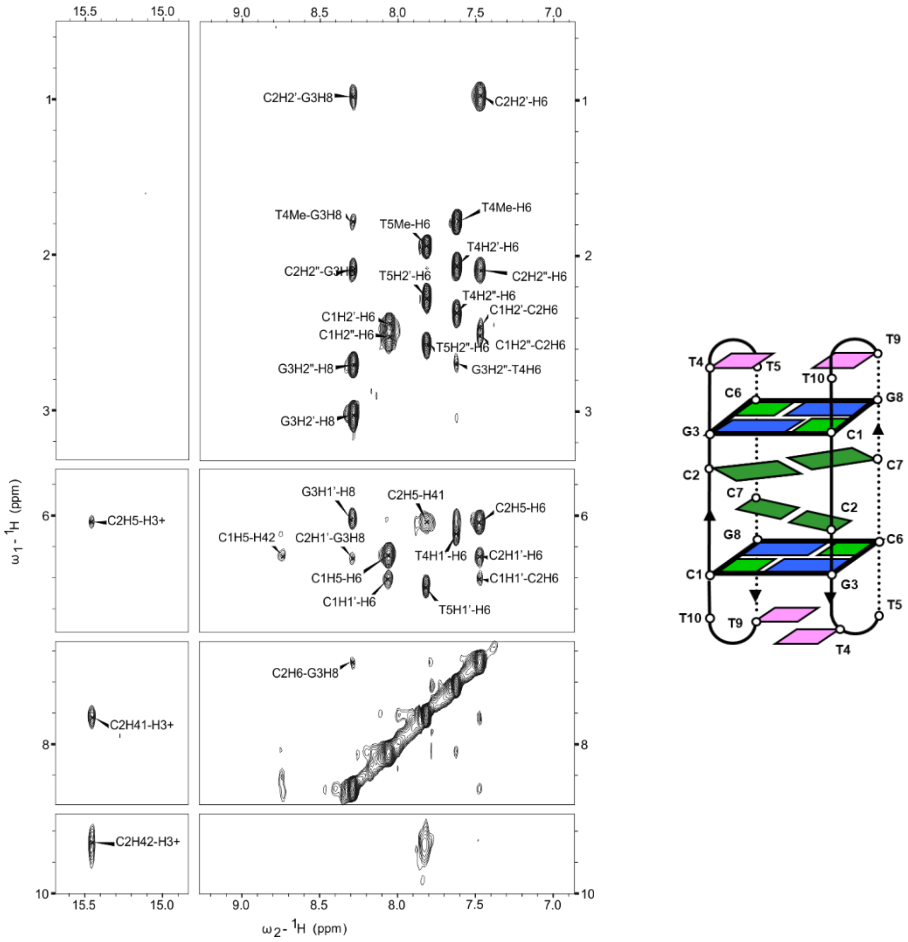


Figure 4-35. Exchangeable and non-exchangeable protons region of the NOESY spectra (150 ms) of **cN** at pH 5 and 5 ºC. H_2O/D_2O 90:10, 10 mM phosphate buffer, [oligonucleotide] = 0.8 mM (left). Schematic representation of the structure of **cN** (right).

Interestingly, for the **cN** oligonucleotide at acidic pH, the formation of the alternative structure found for the **N** sequence it is not observed. The conformational restricted nature of the cyclic 10-mer sequence most likely does not offer the required flexibility to adopt an intercalated structure with four hemiprotonated C:C^+ base pairs.

Chemical shifts of **cN** at pH7 and pH 5 are given in *Tables App* 8 and 9, respectively.

4.5. CONCLUSIONS

For all the mini i-motif species studied in this chapter, the favoring capping interactions arise from slipped minor groove tetrads. This suggests that direct minor groove tetrads do not have the proper geometry to interact favorably with $C:C^+$ base pairs and stabilize i-motif structures.

L sequence folds into a major head-to-tail mini i-motif structure at neutral pH stabilized by two slipped G:C:G:T minor groove tetrads which provide, apparently, higher stabilization than one slipped G:C:G:C and one slipped G:T:G:T tetrad. At acidic pH the head-to-head species appears as a minor species. **cL** sequence exhibits the same behavior than its linear analogue, but a higher stability.

N sequence shows multiple species at neutral pH, being at least two of them i-motifs. Besides the different head-to-tail and head-to-head mini i-motif topologies that coexist in equilibrium, a pH-dependent equilibrium with a third i-motif structure stabilized by G:G:T triads is observed. Interestingly, this acidic species is not observed for the cyclic analogue **cN**.

K sequence confirms that A:T:A:T tetrads are not compatible with mini i-motif structures and reinforces that only minor groove tetrads are compatible with hemiprotonated base pairs.

The obtained stability data evinces that G:C:G:T and G:C:G:C minor groove tetrads provide an improved stabilizing effect compared to the previously reported G:T:G:T tetrads.

5. MINOR GROOVE TEDRADS IN i-MOTIFS. PART II: MONOMERIC STRUCTURES

5.1. BACKGROUND AND OBJECTIVES

The previous chapter showed that different minor groove tetrads are compatible with the formation of mini i-motif structures. Moreover, the stabilizing interactions provided by this singular structural feature seem to be responsible for the surprising observation of these dimeric mini i-motifs at neutral pH. As it was discussed in the previous chapter, the minor groove tetrads responsible for the formation of the most stable structures were the slipped G:C:G:C and G:C:G:T ones. However, the simultaneous formation of multiple species due to the dimeric nature of the structures, together with their low stability at pH 7, prevented us from carrying out the complete structural determination.

The current chapter is focused on designing oligonucleotide sequences able to fold into unimolecular mini i-motif structures, analogous to the dimeric motifs presented in the previous chapter. This approach was encouraged by the good results presented by G. Serrat in his master's degree project. He observed that connecting the two moieties of the dimeric centromeric i-motif described by Gallego *et. al.* with a poly-T linker, does not impede the formation of the structure, but enhances significantly its stability.[223] There are significant differences between the mini i-motif and the centromeric i-motif, comprising the latter a larger $C:C^+$ stack, a unique G:T:G:T minor groove tetrad, and a head-to-head topology. In spite of these differences, we hypothesize that linking the two subunits of dimeric mini i-motifs can give rise to more stable structures without disturbing the general i-motif like fold. Of particular interest is to evaluate the effect provoked by the change in molecularity on the stability of the motif, and on possible topological restrictions triggered by the inclusion of a central linker in the structure (see *Figure* 5-1 for the possible topologies of the folded sequences). Analogously to the dimeric structures, capping interactions by minor groove tetrads are expected to play a key role in the stabilization of the motifs.

5.1.1. REPEAT SEQUENCES STUDIED

The different sets of oligonucleotides studied in this chapter arise from the 9-mer sequences **M**, **L** and **N** presented in previous chapters. In order to adapt these sequences to fold into unimolecular i-motifs instead of their respective dimers, two repeats of the same sequence were joined. The junction was performed by a poly-T linker connecting the 3'-end of one moiety with the 5'-end of the second. The length of the linker was estimated on the basis of the solution dimeric structure of the cyclic analogue of **M** (PDB code 2SLX). Model calculations indicated than a linker of four thymidine residues could be sufficient to connect the two moieties without major distortions. Therefore, the resultant 22-mer sequences used as a starting point for the studies presented in this chapter, respond to the general formula d(*X*-T_4-*X*) and were named as **MM4**, **LL4** and **NN4** (see *Table* 5-1). Based on the conclusions obtained for the dimeric structures on the previous chapter, the hypothetical structures that these sequences will adopt are the ones represented in *Figure* 5-1, expecting the formation of G:C:G:C, G:T:G:T and/or G:C:G:T minor groove tetrads.

Additional sequences were used in this chapter to complete different aspects of our study. For instance, several oligonucleotides in which particular cytosines were substituted by 5-methylcytosine (^{m}C) were utilized to facilitate the assignment of the NOESY spectra. Moreover, a set of oligonucleotides based on **LL4** with different number of thymines in the linker was studied to explore the effect of the linker in the formation and stability of the motif. Finally, longer repetitive oligonucleotide sequences were also studied with the aim of investigating the formation of these sequences in tandem. Each set of sequences will be presented in the corresponding section. All the synthesis studied in this chapter have been purchased from IDT.

NAME	FORMULA	SEQUENCE	EXPECTED TETRADS
MM4	M-T_4-M	d(TCGTTTCGT-TTTT-TCGTTTCGT)	G:T:G:T
LL4	L-T_4-L	d(TCGTTCCGT-TTTT-TCGTTCCGT)	G:C:G:T or G:T:G:T & G:C:G:C
NN4	N-T_4-N	d(CCGTTCCGT-TTTT-CCGTTCCGT)	G:C:G:C

Table 5-1. Sequences of the main oligonucleotides studied in this chapter.

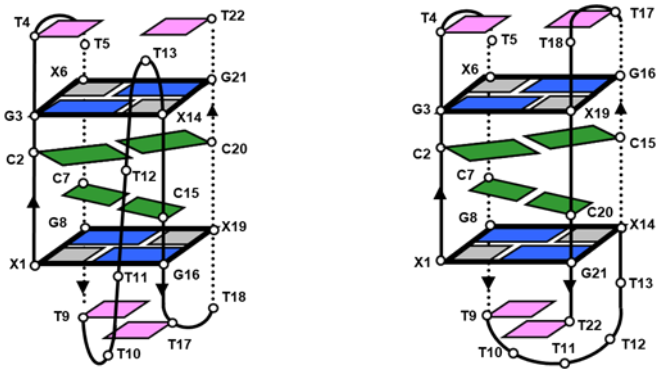


Figure 5-1. Schematic representations of the two possible topologies expected for the sequences studied in this chapter: head-to-tail (left) and head-to-head (right). Cytosine residues are shown in green, guanines in blue and thymines in pink. X residues in grey indicate the variable positions depending on the sequence.

5.2. FORMATION AND STABILITY OF MONOMERIC MINI i-MOTIFS

This section includes the analysis by NMR, CD and UV spectroscopic techniques of different oligonucleotides containing two repeats of M, L, and N sequences. We will study their propensity to form i-motif structures under different temperature and pH conditions as well as their stability.

5.2.1. i-MOTIF FORMATION ASSESSED BY 1D ^{1}H-NMR

As it can be observed in the 1D ^{1}H-NMR spectra recorded at neutral pH shown in *Figure* 5-2, **MM4**, **LL4** and **NN4** sequences are structured at room temperature. In all cases, characteristic signals of protonated N3 nitrogen atoms involved in the formation of C:C^{+} base pairs at around 15 ppm confirm the formation of i-motif-like structures. Analogously to that we observed in the previous chapter, the structures adopted by these oligonucleotides afford very similar NMR signal patterns, indicating that they share a similar folding. Although 2D experiments are necessary to fully characterize the structures, interesting features can be directly spotted from the 1D spectra.

Unexpected guanine imino signals typical of WC G:C base pairs are observed in the 1D ^{1}H-NMR spectrum of **MM4** at pH 7 and 5 ºC. These signals are not compatible with the hypothesized i-motif-like structure for this sequence. These signals, which are overlapped and poorly defined, disappear at higher temperatures and more acidic conditions (*Figure* 5-2), while the rest of the signals remain unaltered at higher temperature. In addition, their intensities are reduced when the experiments are conducted at lower oligonucleotide concentrations. For these reasons we conclude these signals belong to some sort of minor unstable multimeric aggregation, probably a partially self-complementary duplex form.

Considering now the major i-motif species, two imino signals at ~15.5 ppm confirm the formation of hemiprotonated C:C^{+} base pairs. As occur in the case of its dimeric linear and cyclic analogues, G:T:G:T minor groove tetrads are expected to cap the mini i-motif. Characteristic G:T thymine imino signals at 12 ppm support this hypothesis, which must be confirmed by 2D NMR experiments. The numerous signals shifted below 12 ppm belong to imino protons from guanine residues involved in G:T base pairs, bonded amino protons from C:C^{+} base pairs, and probably to amino protons from guanine residues. Although many assignments of all residues are still missing, the formation of a unimolecular mini i-motif stabilized by G:T:G:T minor groove tetrads seems plausible. Based on the NMR melting experiments at pH 7, **MM4** i-motif structure unfolds at temperature values higher than 25ºC.

Based on the exchangeable proton region of the spectra of **LL4**, the simultaneous formation of C:C^{+} and WC G:C base pairs can be unequivocally confirmed. As mentioned above, the characteristic imino protons of the hemiprotonated base pairs appear downshifted above 15 ppm. Two imino protons of guanine residues forming canonical G:C pairs are found between 13 and 14 ppm. Moreover, several signals between 11 and 12 ppm potentially assignable to thymine imino protons in G:T mismatches are also observed. This preliminary information points to the formation of mini i-motif structures stabilized by either G:C:G:T or G:C:G:C and G:T:G:T minor groove tetrads. According to the number and pattern of the signals in the exchangeable proton region, a unique major species seems to be predominant under these experimental conditions for **LL4**. Of special interest will be to compare the topology of this major species to the one adopted by the dimeric structure of **L**, which will give precious information on the influence provided by the type of tetrad formed and the presence of the poly-T linker. More signals comprised between 9 and 12 ppm complete the set of exchangeable protons of the base pairs described

above (H1G, H22G and H42C). Most of the exchangeable protons are still observed in the NMR spectra recorded at 25ºC and 35ºC, indicating the structure is more stable than the dimeric analogue. Interestingly, all exchangeable signals disappear concomitantly upon increasing the temperature, suggesting a cooperative unfolding.

Finally, **NN4** also shows the formation of an i-motif-like structure at neutral pH very clearly. Interestingly, together with the major species at pH 7, tiny imino signals in the hemiprotonated C:C$^+$ region suggest the coexistence of minor i-motif species in equilibrium. The discussion of the structure of this minor species will be performed in following sections. Focused on the major i-motif species, WC G:C base pairs formation is corroborated by the signal at ~13.4 ppm attributable to base paired guanine imino protons. Other two imino signals appear at around 12 ppm. Considering the sequence of **NN4**, no G:T mismatches are expected to be formed if the oligonucleotide folds into a mini i-motif structure stabilized by G:C:G:C minor groove tetrads. However, the chemical shifts of these signals do not match canonical WC G:C base pairing. Thus, the nature of these signals will need further investigation to be determined. As for the other sequences, amino protons from guanine and/or cytosine residues appear between 9 and 10 ppm. **NN4** i-motif structure presents a similar stability than that of **LL4**, observing residual signals from the major species up to 35 ºC.

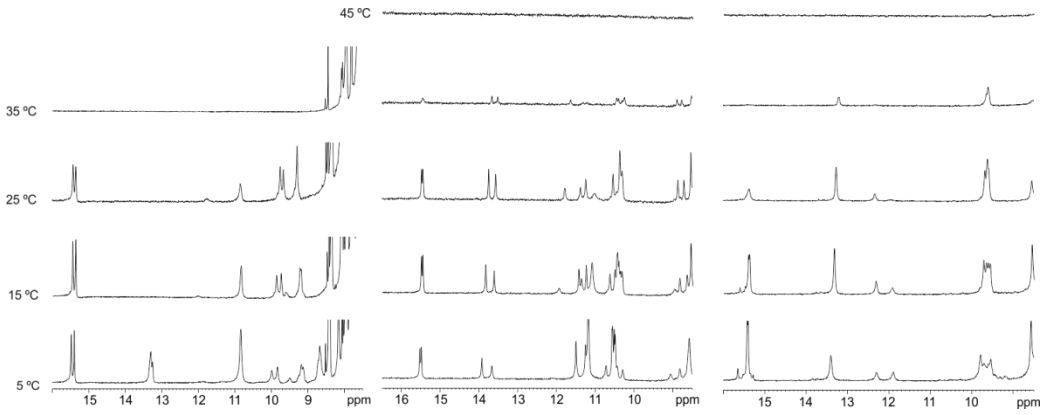


Figure 5-2. 1D ^{1}H-NMR melting series of **MM4** (left), **LL4** (center) and **NN4** (right) at pH 7. 10 mM phosphate buffer (H_2O:D_2O 90:10), [oligonucleotide] = 1 mM.

5.2.2. STRUCTURES ARE MONOMERIC

In order to confirm that **MM4**, **LL4** and **NN4** sequences fold into monomeric species, NMR experiments performed at different oligonucleotide concentration were recorded. The results obtained for the concentrated and the diluted samples are shown in *Figure* 5-3 for comparison.

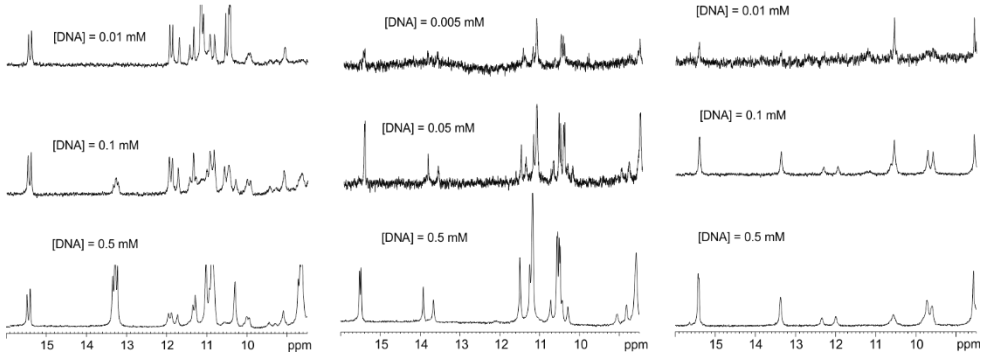


Figure 5-3. 1D ^{1}H-NMR spectra of **MM4** (left), **LL4** (middle) and **NN4** (right) at different oligonucleotide concentration.10 mM phosphate buffer (H_2O:D_2O 90:10), pH 7, T=5 ºC.

As it can be observed, no significant changes are found for any sequence upon dilution on the exchangeable proton signals of the spectra. The same number of signals at the same chemical shift with analogous relative intensity allows concluding that the same species are formed in both conditions. Therefore, as it is characteristic for unimolecular constructions, the currently studied oligonucleotides fold into i-motif structures in a concentration-independent manner.

5.2.3. EXPERIMENTS AT ACIDIC pH REVEAL DIFFERENT BEHAVIOR OF THE SEQUENCES

Given that i-motif structures are especially sensitive to pH modulations, their study under acidic conditions might reveal important information. Furthermore, assuming the hypothetical formation of a unimolecular mini i-motif for the **MM4** sequence will include the assembling of G:T mismatches, we expect this sequence to remain folded in a wide range of pH. On the other hand, for **LL4** and **NN4** sequences, the presence of WC G:C makes them more susceptible to structural changes due to the protonation of cytidine residues, which raises the interest of studying these sequences under different pH conditions.

Figure 5-4 shows the 1D ^{1}H-NMR spectra of **MM4**, **LL4** and **NN4** at 5ºC and different pH values. Notably, the signals corresponding to the major species of **MM4** and **LL4** do not suffer significant alterations upon acidification. As mentioned in the previous section for the **MM4** , the imino signals of WC G:C base pairs, which we assumed to belong to a minor non-i-motif species, become less intense at pH 6 and totally disappear at pH 5. Significantly, the pattern of the rest of the signals remains unaltered under these acidic conditions, showing only an increase in their intensity and a better dispersion, probably related to a greater stability of the motif. Very interestingly, the same behavior is observed for **LL4**. This information contrasts with what we observed for its shorter analogue **L**, which showed the coexistence of two species corresponding to head-to-head and head-to-tail topologies at acidic pH values. The

signals corresponding to the formation of WC G:C base pairs are not observable at pH 5. Signals below 12 ppm become narrower and more intense at acidic pH.

Contrary to its related sequences, **NN4** 1D ^{1}H-NMR spectra under acidic conditions show important changes compared to the analogous experiment at neutral pH. The most intense signal at ~15.4 ppm at neutral pH considerable diminishes its intensity at pH 5, while the tiny signals at ~15.3 and ~15.6 ppm at neutral pH become predominant upon acidic conditions. Furthermore, the imino protons signals from guanine residues forming WC G:C and the two undetermined signals observed at ~12 ppm completely disappear from pH 7 to pH 6. Additionally, a relevant set of signals appears at acidic pH in the imino-amino protons region (10-12 ppm) matching the theoretical formation of G:T mismatches predicted for **MM4** and **LL4**. All these changes in the spectra suggest that **NN4** presents a pH-dependent equilibrium between different i-motif species, in a similar way as observed for oligonucleotide **N**.

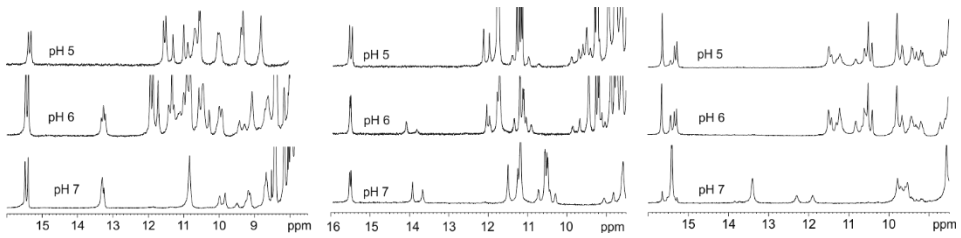


Figure 5-4. 1D ^{1}H-NMR series of **MM4** (left), **LL4** (center) and **NN4** (right) at 5 ºC and different pH values. 10 mM phosphate buffer (H_2O:D_2O 90:10), [oligonucleotide] = 1 mM.

In order to support the 1D ^{1}H-NMR data previously presented, CD spectra were recorded at different pH values. *Figure* 5-5 shows that the CD spectra of the three oligonucleotides respond similarly to pH variations. The spectroscopic features of the three folded sequences afford CD spectra with maximum bands between 260 and 270 nm and minimum bands around 240 nm. These profiles match the ones observed for other mini i-motif structures and differ from other more standard i-motifs.[76] Upon basification, the intensity of the bands is reduced, and the maximum is shifted to higher wavelength values, responding to a random coil profile. This transition clearly occurs above pH 7 for **LL4** and **NN4** and for lower pH values in the case of **MM4**, which will be quantitively analyzed in the following section. Very interestingly, a less intense maximum band is observed for the **NN4** sequence at acidic pH values. Remarkably, this maximum band matches to the formation of a more standard i-motif structure (295 nm) and disappears at neutral conditions. This information confirms the formation of an alternative i-motif structure under acidic conditions for **NN4**.

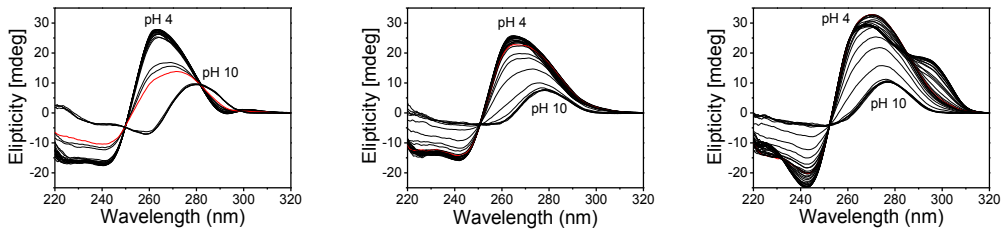


Figure 5-5. CD spectra superposition of **MM4** (left), **LL4** (center) and **NN4** (right) at 5 ºC and different pH values. 25 mM phosphate buffer, [oligonucleotide] = 3.5 µM. Spectra at neutral pH are shown in red.

As a summary of the conclusions outlined in this section, **MM4**, **LL4** and **NN4** sequences fold into major i-motif species at neutral pH. While **LL4** exclusively shows the formation of one major species from acidic to basic pH, **MM4** and **NN4** present alternative constructions. **MM4** i-motif coexists with a duplex or multimeric aggregation involving G:C WC base pairs at low temperatures and neutral pH, and **NN4** shows a pH-dependent equilibrium between different i-motif structures.

5.2.4. THE STABILITY OF THE STRUCTURES IS pH-DEPENDENT

In this section we quantitively study the formation of the structures as a function of pH and temperature. The obtained data is summarized in *Table* 5-2.

NAME	FORMULA	pH_T	T_m pH 7	T_m pH 6.0	T_m pH 5.0
MM4	M-T_4-M	6.7 ± 0.1	16.6 ± 0.5	29.5 ± 0.5	38.0 ± 0.5
LL4	L-T_4-L	7.9 ± 0.1	30.5 ± 0.5	44.3 ± 0.5	50.3 ± 0.5
NN4	N-T_4-N	8.0 ± 0.1	29.2 ± 0.5	43.4 ± 0.5	53.3 ± 0.5

Table 5-2. Summary of the stability results (pH_T and T_m values) obtained for **MM4**, **LL4** and **NN4**. 25 mM phosphate buffer, [oligonucleotide] = 2 µM. pHT values were obtained from CD pH titration (see Figure 5-5) and T_m were calculated from UV melting experiments.

From the CD experiments showed in *Figure* 5-5, pH titration curves were obtained by plotting the maximum ellipticity values at specific wavelengths (*Figure* 5-6). All three curves show a similar behavior, except **NN4**. Describing the graphs from acidic to basic pH values, ellipticity slightly increases at pH values close to the pK_a of the nucleobases due to their deprotonation. From this point onwards, a plateau is drawn indicating the formation of a stable structure throughout a range of pH. While in the case of **MM4** and **LL4** the plateau of high ellipticity leads to a drop in ellipticity due to the unfolding of the structure, **NN4** shows an initial transition from high to higher ellipticity values between pH 5 and pH 7

that corresponds to the equilibrium previously described between the alternative i-motif species. Further explanations will be given in Chapter 6 regarding the two transitions observed for **NN4**. The transition to the random coil at neutral or even basic pH draws a sigmoidal curve that can be statistically adjusted to a Boltzmann fit in order to obtain pH_T values for the three sequences. Remarkably, **LL4** and **NN4** pH_T values are notably higher (Table 5-2) than most of the reported examples of highly stable i-motif structures.[74,75,197] On the other hand, **MM4** unfolding takes place closer to neutral pH conditions. Qualitatively, the pH_T values of these three repetitive constructions endorse the results obtained in the previous chapter for the qualitative ranking of the stabilization provided by the different minor groove tetrads.

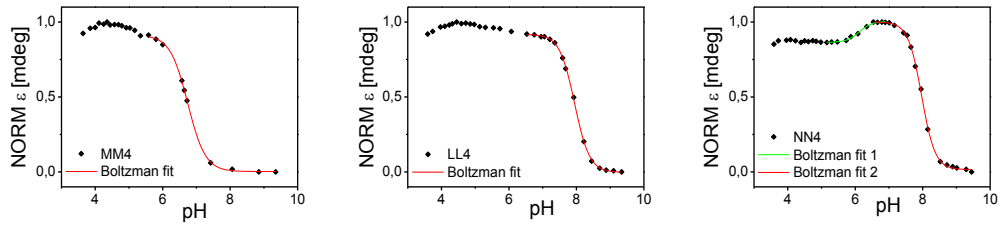


Figure 5-6. CD-monitored pH titration curves of **MM4** (left), **LL4** (center) and **NN4** (right) at 5 ºC. 25 mM phosphate buffer, [oligonucleotide] = 2.0-3.5 µM. Boltzmann fit represented in red for the denaturalization of the structures and in green for the equilibrium between species of **NN4**.

The thermal stability of these sequences was evaluated by comparison of their T_m values at different pH conditions (Table 5-2). UV melting experiments following the unfolding of the previously annealed structures at 260 nm were performed for the obtention of this data, which resulted in nicely defined sigmoidal curves that were then adjusted to a Boltzmann fit (see *Figure* 5-7). As expected, all sequences significantly increase their stability at acidic pH values. Notably, such increase does not follow a linear trend, being the difference of T_m values for the three sequences more remarkable from pH 7 to pH 6 than from pH 6 to pH 5. This fact suggests that stronger acidic conditions will not enhance the stability of the structures. As it was introduced from the NMR experiments at different temperatures, **MM4** exhibits a low stability at neutral pH (16.6ºC). This lower stability compared to the other sequences is also observed under acidic conditions, showing T_m values 10 to 15 ºC lower than those of **LL4** and **NN4** in all cases. **LL4** and **NN4** show very similar stability data (T_m ~30 ºC, pH 7). Interestingly, while at pH 7 and pH 6 **LL4** is approximately 1ºC more stable than **NN4**, at pH 5 **NN4** is more stable than **LL4**. This change in the relative stability of **LL4** and **NN4** may be related to the presence of an alternative species for **NN4**. As in the case of the pH_T studies, the results obtained for **MM4**, **LL4** and **NN4** are consistent with those obtained for **M**, **L** and **N**, which support that the most stable mini i-motif structures are those capped by G:C:G:T and G:C:G:C minor groove tetrads.

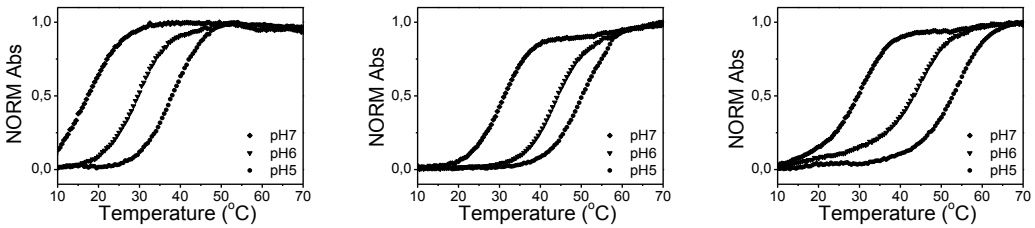


Figure 5-7. Superposed UV-monitored melting curves of **MM4** (left), **LL4** (center) and **NN4** (right) at different pH values. 25 mM phosphate buffer, [oligonucleotide] = 2.0-3.5 mM.

5.2.5. THE LINKER AFFECTS THE STABILITY OF THE STRUCTURES BUT NOT ITS FORMATION

Since the construction of the linker with four nucleotides was established on a qualitative approximation, we wanted to assess if different linker lengths are also capable of affording i-motif formation. For this experiment, a set of oligonucleotides derived from **LL4** containing poly-T linkers from one to seven residues (**LL1-7**) was studied. If the i-motif formation is feasible for other linker lengths, of special interest will be to evaluate any changes in topology or stability provoked by the linker modification.

Remarkably, all the recorded 1D ^{1}H-NMR and CD spectra at neutral pH for the sequences **LL1-7** exhibit very similar features (*Figure* 5-8): NMR signals corresponding to the formation of hemiprotonated C:C^{+} base pairs, together with the simultaneous observation of signals matching the formation of G:C and G:T base pairs. Also, the CD spectra of these sequences maintain the characteristic positive band around 265 nm previously described for **LL4**. These results not only settle that both shorter and longer linkers are compatible with the formation of i-motif structures for these set of repetitive constructions, but also that the main structural features of these motifs are formed independently of the length of the linker.

Next, the thermal stability of these structures was evaluated by CD melting experiments. The obtained melting curves are presented overlapped in *Figure* 5-9. Corresponding T_m values were calculated using the Boltzmann fitting method(ref) and are summarized in *Table* 5-3. Interestingly, **LL4** shows one of the greatest thermal stabilities (28.4 ºC), only improved by **LL3** (32.1 ºC). The less stable structure is the one adopted by the longest studied sequence (**LL7**, 21.7 ºC). Therefore, all T_m values for these set of sequences are comprised in a range of ~10 ºC.

Given the great thermal stability showed by **LL3**, its pH_T value was calculated from pH titration experiments monitored by CD spectroscopy (*Figure* 5-9*, right)*. The obtained value (7.8) is again remarkably high for an i-motif structure and very similar to that of **LL4** (7.9), which difference lays within the experimental error.

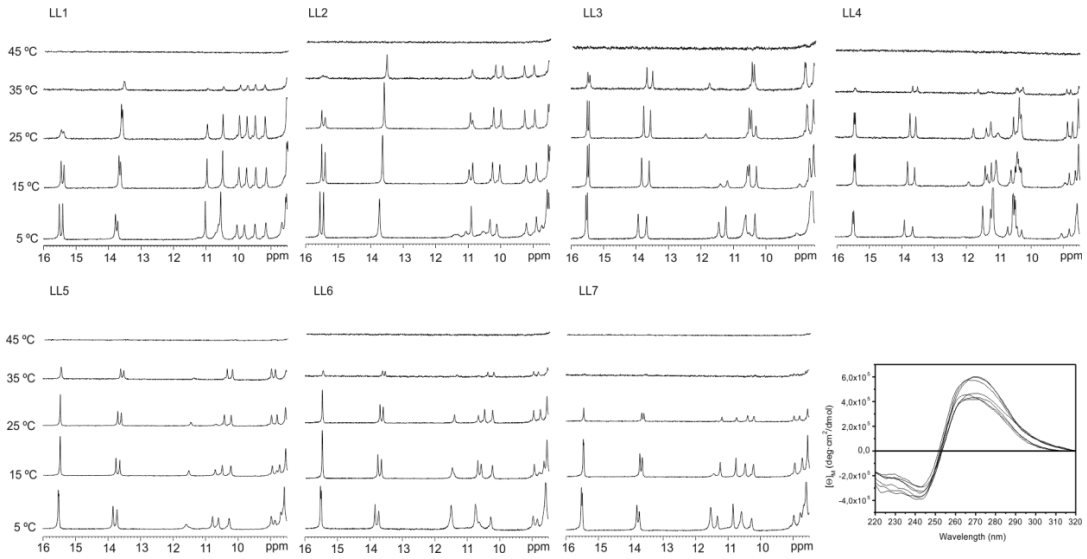


Figure 5-8. 1D ^{1}H-NMR melting series of **LL1-7** at pH 7. 10 mM phosphate buffer, [oligonucleotide] = 1 mM. Bottom-right corner: superposed CD spectra of **LL1-7** at 10 ºC and pH 7. 25 mM phosphate buffer (H_2O:D_2O 90:10), [oligonucleotide] = 3.5 µM.

NAME	SEQUENCE	pH_T	T_m pH 7
LL1	L-T_1-L	nd	24.3 ± 0.5
LL2	L-T_2-L	nd	27.9 ± 0.5
LL3	L-T_3-L	7.8 ± 0.1	32.1 ± 0.5
LL4	L-T_4-L	7.9 ± 0.1	28.4 ± 0.5
LL5	L-T_5-L	Nd	25.4 ± 0.5
LL6	L-T_6-L	nd	23.3 ± 0.5
LL7	L-T_7-L	nd	21.7 ± 0.5

Table 5-3. Summary of the stability results (pH_T and T_m values) obtained for **LL1-7** sequences. 25 mM phosphate buffer, [oligonucleotide] = 3.5 µM. nd: not determined

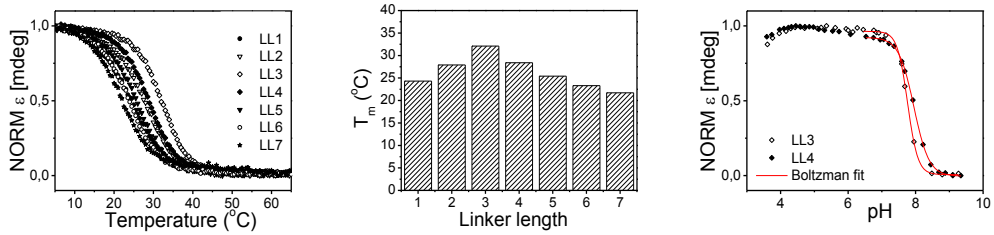


Figure 5-9. Superimposed DC-monitored melting curves of **LL1-7** at pH 7 (left). Graphic representation of T_m values calculated for **LL1-7** melting curves at pH 7 (center). Superimposed CD-monitored pH titration curves of **LL3** and **LL4** at 5 ºC. (right). Boltzmann fit represented in red for the denaturalization of the structures.

5.2.5.1. **LL1-7** STRUCTURES ARE MONOMERIC

In order to ensure that all **LL1-7** sequences fold into monomeric species, native electrophoretic (PAGE) experiments were carried out to assess the mobility of the structures compared to reference oligonucleotides of different length (*Section* 8.8 in Methods). Shown in *Figure* 5-10, the comparison of the mobility of the **LL1-7** oligonucleotides (19-25mers) with a longer oligonucleotide (**LL3rep**, 48mer) and poly-T ladders of different length under native conditions confirms the unimolecularity of the structures. Using the poly-T ladders as references, it is evident that all **LL1-7** sequences migrate between 15 and 30mers, concordantly showing a reduced mobility from **LL1** to **LL7**. As a further confirmation, **LL3rep** matches the migration of a 50mer poly-T.

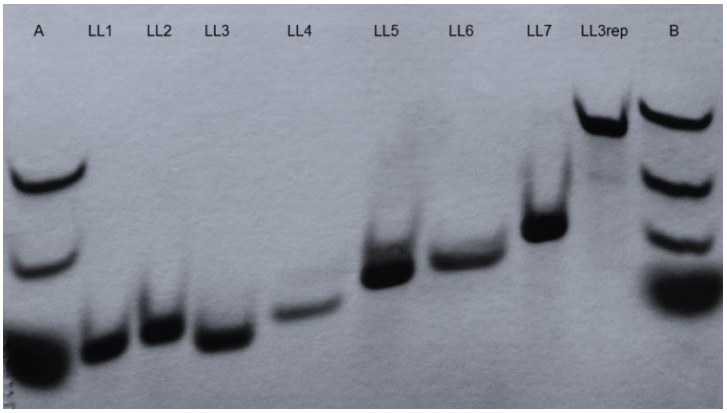


Figure 5-10. Native PAGE of **LL1-7**. 20% polyacrylamide gel (acrylamide/bisacrylamide 19:1), pH 7, phosphate buffer 0.05 M. Experimental conditions: 180 V for 6-7 hours, T=5°C. A = dT15, dT25, dT40 + bromophenol. B = dT20, dT30, dT50 + bromophenol.

5.2.5.2. THERMODYNAMICS

As it was introduced in the *Section* 2.2.4 in Fundamentals, thermodynamic parameters can be extracted from melting curves. The case of the **LL1-7** sequences is especially suitable for the thermodynamic study as, contrary to **MM4** and **NN4**, fold into single unimolecular structures, and a two-state equilibrium between folded and unfolded species can be assumed. Therefore, knowing the molecularity, single melting experiments have been used to obtain equilibrium-affinity constant (*K*) values over a wide range of temperatures.[171] Van't Hoff plots (ln*K* vs 1/T) for each sequence are shown in *Figure* 8-1 in Methods. The analysis of the temperature range was restricted to the region $0.15<x<0.85$, in which a linear behavior was obtained. The thermodynamic parameters from the folding equilibrium of the **LL1-7** oligonucleotides at pH 7 are presented in *Table* 5-4.

SEQUENCE	T_m (°C) (Boltz.)	THERMODYNAMICS (Van't Hoff)				
		0.85 > x > 0.15				
		Equation	ΔH^0 (KJ/mol)	ΔS^0 (J/K·mol)	ΔG^0_{278} (KJ/mol)	$T_m = \Delta H^0 / \Delta S^0$ (°C)
LL1	24.3 ± 0.5	y=18602x-62.554 R^2=0.9974	-155 ± 1	-520 ± 5	-10 ± 3	24.2
LL2	27.9 ± 0.5	y=20764x-69.073 R^2=0.9942	-173 ± 3	-575 ± 8	-13 ± 5	27.5
LL3	32.1 ± 0.5	y=25340x-83.061 R^2=0.9946	-211 ± 3	-691 ± 10	-19 ± 6	31.9
LL4	28.4 ± 0.5	y=23115x-76.613 R^2=0.9968	-192 ± 2	-637 ± 7	-15 ± 4	28.6
LL5	25.4 ± 0.5	y=21593x-72.298 R^2=0.9933	-180 ± 3	-602 ± 10	-13 ± 6	25.5
LL6	23.3 ± 0.5	y=21248x-71.750 R^2=0.9976	-177 ± 2	-596 ± 6	-12 ± 3	23.7
LL7	21.7 ± 0.5	y=20197x-68.403 R^2=0.9974	-168 ± 1	-569 ± 5	-10 ± 3	22.1

Table 5-4. Thermodynamic parameters calculated from the melting curves of **LL1-7** at pH 7.

T_m values obtained from the Van't Hoff plot match the ones calculated on the basis of the Boltzmann equation. This match gives confidence in the quality of the fits. Melting curves of all the sequences exhibit similar shapes so, as expected, more favorable Gibbs free energy values are obtained for the sequences that exhibit higher T_m values, LL3 and LL4. Enthalpy values are in concordance to Gibbs free energy values, indicating the importance of the enthalpic contribution to the stability of the structures. Regarding entropy values, one may expect sequences with longer loops to have more unfavorable entropic contributions. However, this tendency is not observed. Longer loops may be partially ordered and changes in water solvation may also have a great effect. Enthalpy-entropy compensations have been described to occur in DNA folding/unfolding processes, because of solvent reorganization, but also because favorable enthalpic interactions generally imply the loss of degrees of freedom. In this case, the most unfavorable entropic contributions correspond to LL3 and LL4 with the best enthalpic values.

5.3. STRUCTURAL CHARACTERIZATION

To confirm that **MM4**, **LL4** and **NN4** fold into mini i-motif structures, a complete characterization by 2D NMR spectra was carried out. Important questions remained unanswered, such as the nature of the minor groove tetrads or the topology of the structures. Due to the repetitive and highly symmetric nature of the sequences, the incorporation of 5-methylcytosines (^{m}C) was required for the specific assignment of residues located in similar chemical environments. As it was mentioned in previous sections, we put

more attention in **LL4** and related sequences since their sequences are slightly less repetitive than **NN4** and **MM4** and, therefore, more convenient for NMR studies.

5.3.1. MM4 FOLDS INTO AN I-MOTIF STABILIZED BY G:T:G:T MGT

Characterization of **MM4** required 2D spectra to be recorded at pH 5 and 5ºC, conditions at which WC-like alternative species is not formed. Moreover, due to the lower stability of **MM4** compared to that of **LL4** and **NN4**, the quality of the NOESY spectra is improved at acidic pH, simplifying its assignment.

As mentioned, in order to overcome the high overlapping of the NOE signals, specially observed in the exchangeable proton region, the study of ^{m}C-containing sequences was carried out. **MM4_^{m}C(2)** and **MM4_^{m}C(2,15)**, which incorporate 5-methylcytosine residues substituting cytosines in positions 2 or 2 and 15 of **MM4**, respectively, were designed to enable the specific assignment of the sequence.

Rather unexpectedly, **MM4_^{m}C(2)** and **MM4_^{m}C(2,15)** structures afforded very low-quality NOESY and, moreover, show residual cross-peaks belonging to the WC-like alternative species. This fact impeded the full assignment of these sequences, whereas analogous strategies resulted very successful in the case of **LL4** and **NN4** (see 5.3.2.1). Nonetheless, some crucial cross-peaks from the NOESY and TOCSY spectra of **MM4_^{m}C(2)** were identified and allowed the identification of the spin systems belonging to ^{m}C2 and C15 residues, which were then identified in the NOE spectra of **MM4** by analogous chemical shifts of H5-H6 TOCSY signals.

Now, from these unequivocal starting points, the characterization of the main structural features of **MM4** at pH 5 could be achieved. The great number of exchangeable-protons signals of the NOESY spectra (*Figure* 5-11) allowed the confirmation of a head-to-tail orientation for **MM4**, exhibiting as main stabilizing units two hemiprotonated C:C^{+} base pairs capped by two G:T:G:T minor groove tetrads. The general workflow for the specific characterization of the cross-peaks was the following: From H5 protons of C2 and C15 residues, stacking cross-peaks with the imino protons of G3:T14 (9.4 and 11.4 ppm) and G16:T1(11.5 and 11.7 ppm) mismatches were respectively identified. The H1G protons of these mismatches show cross-peaks with H1'G protons of other guanine residues that were identified as H1G3-H1'G21 and H1G16-H1'G8. This information unequivocally confirms the head-to-tail topology of the motif (*Figure* 5-12) and allows to characterize, after identifying H1G8/G21-H3T14/T19 cross-peaks (11.0/10.9 and 11.9/12.0 ppm, respectively), the formation of two perfectly defined G:T:G:T minor groove tetrads. The formation of these tetrads is also supported by inter-strand H3T-H2'/H2"T cross-peaks across the major groove. Finally, a great number of stacking contacts between H3^{+}C and H1/H22/H21G atoms allowed the characterization of the four 5'-CG-3' steps (C2-G3, C7-G8, C15-G6 and C20-G21). Then, imino-amino cross-peaks served for the characterization of the hemiprotonated base pairs C2:C20^{+} (H3^{+}: 15.4 ppm) and C7:C15^{+} (H3^{+}: 15.5 ppm).

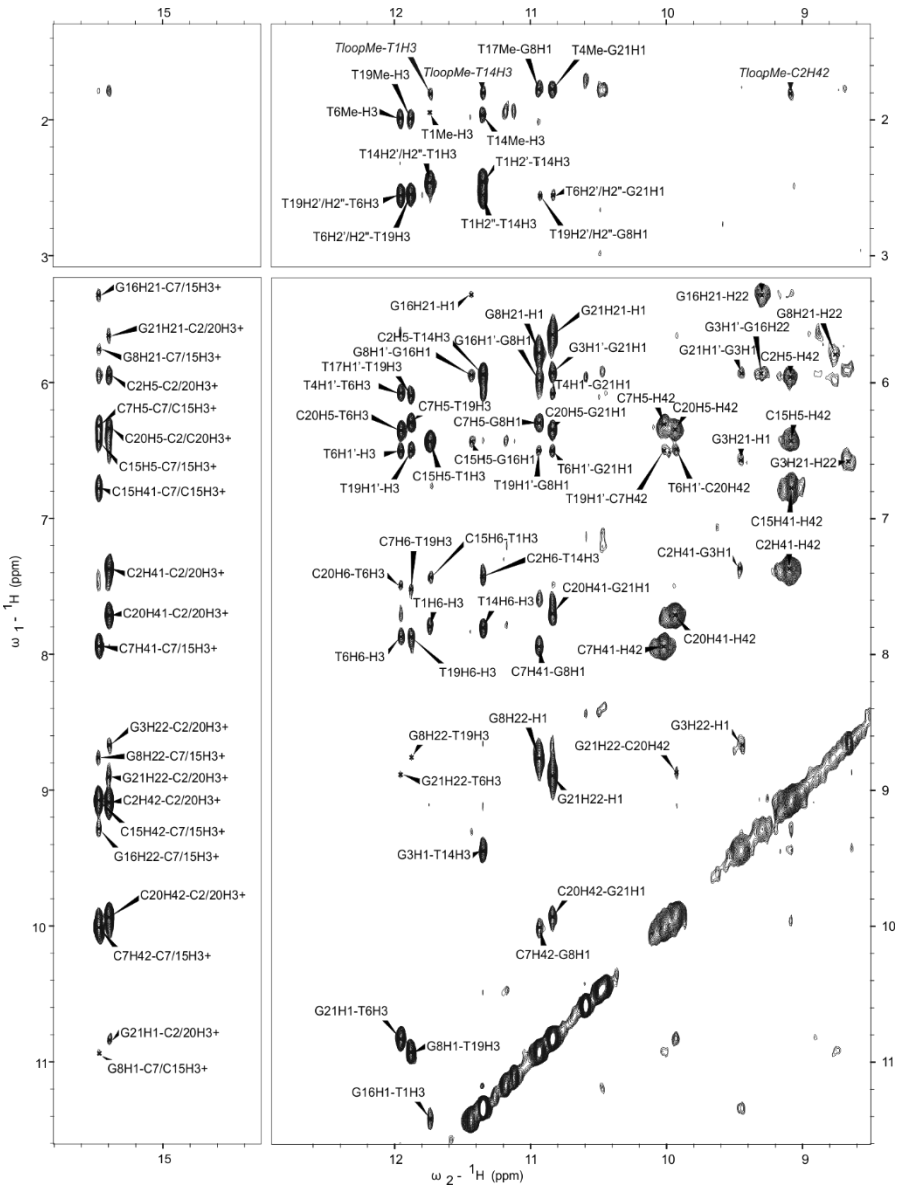


Figure 5-11. Exchangeable protons regions of NOESY spectrum (150 ms) of **MM4** at pH 5 and 5 ºC. H_2O/D_2O 90:10, 10 mM phosphate buffer, [oligonucleotide] = 1 mM.

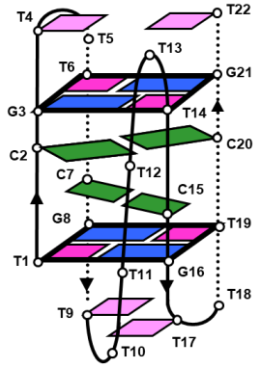


Figure 5-12. Schematic representation of the structure of **MM4** at pH 5.

5.3.2. LL3 AND LL4 FOLD INTO AN I-MOTIF STABILIZED BY G:T:G:C TETRADS

NOESY spectra of **LL3** and **LL4** recorded at neutral pH exhibit no substantial differences, confirming that their structures are very similar. Both spectra display narrow lines and very good dispersion of the NOE signals, especially in the exchangeable proton region, which facilitates their assignment.

Initially, NOE cross-peaks between sequentially connected 5'-CGT-3' fragments were clearly observed. Considering the two hemiprotonated imino signals at around 15.5 ppm in the NOESY spectra of **LL4** (*Figure* 5-18), each of them shows cross-peaks with two pairs of amino protons, pointing towards the formation of C:C$^+$ base pairs between non-equivalent residues. The pairs of amino protons from cytidine residues involved in the formation of the same C:C$^+$ base pair, show a remarkable different chemical shift. This information, would strongly support the formation of mixed G:C:G:T tetrads, rather than G:C:G:C and G:T:G:T tetrads, as in this case each cytosine would be stacked with guanine residues involved in the formation of different base pairs. This conclusion is confirmed by a number of cross-peaks between amino and H1' protons of different guanines involved in the formation of G:C and G:T base pairs.

However, due to the repetitive nature of the sequences and the symmetric disposition of the structure, the specific assignment of the residues could not be achieved unambigously. In order to overcome this impediment, two sequences containing 5-methylcytosine substituting strategic cytosine residues from **LL4** were used. **LL4_^{m}C(2)** and **LL4_^{m}C(2,15)** incorporate one ^{m}C residue in position 2, and two ^{m}C residues in positions 2 and 15, respectively.

Full details and specific assignments are detailed in the following sections.

5.3.2.1. **LL4_^{m}C(2)** AND **LL4_^{m}C(2,15)** ENABLE THE SPECIFIC ASSIGNMENT OF **LL4**

As it can be observed from the NMR spectra shown in *Figure* 12, the incorporation of a single or two methylated residues in **LL4** does not significantly alter the formation nor the stability of the i-motif structure at neutral pH.

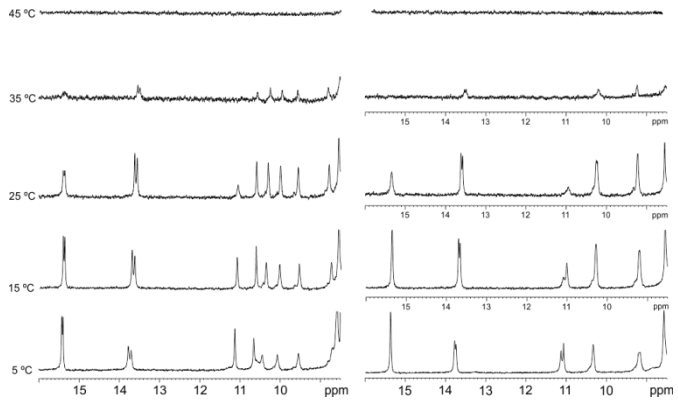


Figure 5-13. NMR melting series of **LL4_^{m}C(2)** (left) and **LL4_^{m}C(2,15)** (right) at pH 7. 10 mM phosphate buffer (H_2O:D_2O 90:10), [oligonucleotide] = 0.5 and 1 mM for **LL4_^{m}C(2)** and **LL4_^{m}C(2,15)**, respectively.

Moreover, natural cytosines H5 and H6 protons in modified sequences exhibit similar chemical shifts than that found in the unmodified sequences. This is important, because, once assigned **LL4_^{m}C(2)** and **LL4_^{m}C(2,15)**, the comparison of the H5-H6 protons region of the TOCSY allowed the specific assignment of **LL4** and **LL3** sequences, as it will be later discussed.

Exchangeable protons region of the 2D NOESY spectrum of **LL4_^{m}C(2,15)** is especially relevant for determining the folding topology of the structure. This is, cytosines in position 2 and 15 would be involved in the formation of the same C:C$^+$ base pair in the head-to-head like structure, but they would participate, in different C:C$^+$ base pairs in a head-to-tail orientation of the structure (see Figure 5-1)

As shown in *Figure* 5-14, the NOESY spectra recorded at pH 7 for **LL4_^{m}C(2,15)** shows a high number of degenerated signals, indicating the formation of a remarkably symmetric structure. This elevated structural symmetry is specially reflected in the TOCSY spectrum (*Figure* 5-17*, center right*) which only shows two H5-H6 cross-peaks belonging to overlapped C6/C19 and C7/C20 residues. According to aromatic signals, only 15 spin systems were identified. Although some residual signals were also observed, the most intense spin systems corresponded to a unique structured i-motif species. ^{m}C2 and ^{m}C15 residues (labeled as Cma and Cmb in the spectrum showed in *Figure* 5-14) were identified based on their amino protons, which lack H5 proton connections but show NOE signals with their Me protons. Although the H42 protons of these residues are practically degenerated (9.20 ppm), their H41 protons show enough dispersion (7.62/7.42 ppm) to conclude, through their NOE connections with imino protons

from hemiprotonated C:C$^+$ base pairs (15.36/15.38 ppm), that the structure of **LL4_^{m}C(2,15)** exhibits two ^{m}C:C$^+$ base pairs, only compatible with a head-to-tail topology. Obviously, these imino protons from the ^{m}C:C$^+$ base pairs also show NOE cross-peaks with the amino protons of C7 and C20, which are degenerated. Unfortunately, the great overlapping of the rest of the NOE cross-peaks prevented the unambiguous assignment of all the residues, which were assigned as undifferentiable pairs (see *Table App* 10). However, the main contacts match the formation of a mini i-motif structure. At 13.76 ppm, it can be observed an imino proton signal that corresponds to two guanine residues (with degenerated exchangeable signals) involved in G:C WC base pairs with C6/C19. This imino signal also shows cross-peaks with amino protons of C7/C20 residues, indicating stacking interactions. At 11.15 and 11.08 ppm, two additional guanine imino signals are observed (labelled as Ga and Gb). These signals exhibit cross-peaks with their own amino protons and with the Me group of methylated cytidines ^{m}C2/^{m}C15, indicating that may correspond to stacked G3 and G16 residues. Although no further cross-peaks allow the complete characterization of G:C:G:T minor groove tetrads, the established topology of the motif, the different chemical shift of guanine imino protons and T4/17Me-G8/21H1 cross-peaks supports a head-to-tail mini i-motif structure.

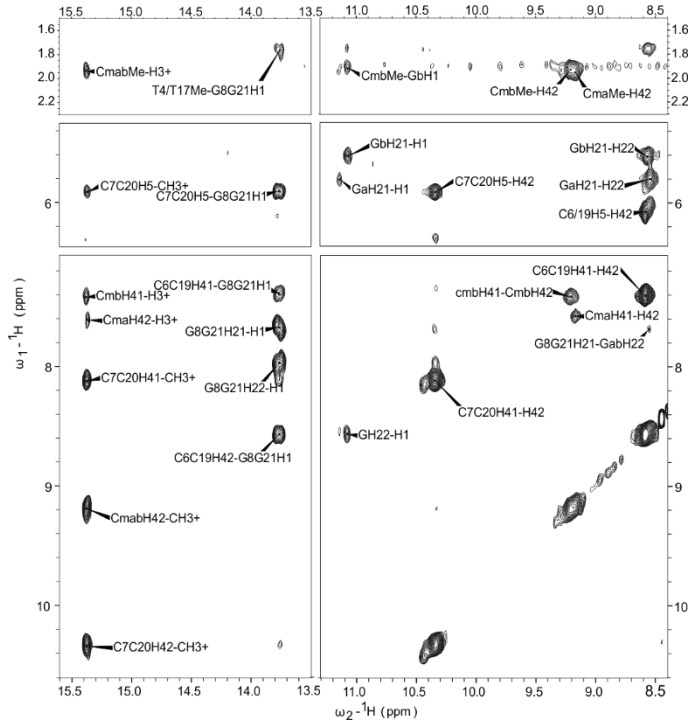


Figure 5-14. Exchangeable protons regions of NOESY spectrum (150 ms) of **LL4_mC(2,15)** at pH 7 and 5 ºC. H_2O/D_2O 90:10, 10 mM phosphate buffer, [oligonucleotide] = 1 mM. Ambiguously assigned residues are labeled with letters.

Similarly, **LL4_^{m}C(2)** spectra exhibits two signals in the characteristic region of hemiprotonated C:C^{+} base pairs (15.44 and 15.41 ppm) that show cross-peaks with two pairs of amino protons, consistent with the formation of C:C^{+} pairs between non-equivalent cytosines. Moreover, WC G:C base pairs formation is supported by imino signals observed at 13.77 and 13.71 ppm that show NOE cross-peaks with amino protons from two cytidine residues. The methylated ^{m}C2 residue (labelled as C2m in the spectrum shown in *Figure* 5-15) is easily recognized by the lack of H5-H6 TOCSY cross-peak and for the numerous Me^{m}C2 contacts. Taking this residue as the starting point, the specific and sequential assignment of the other residues could be accomplished. Given the head-to-tail topology of the structure confirmed by **LL4_^{m}C(2,15),** the other cytosine residues involved in C:C^{+} base pairs could be unambiguously assigned: ^{m}C2 is hydrogen bonded to C15 and the other C:C^{+} base pair is formed between C7 and C20. C6 and C19, that show degenerated signals, are involved in G:C WC base pairs formation (C6:G21 and C19:G8) and expected H1G21-H41/H42C6 and H1G8-H41/H42C19 cross-peaks are easily identified. Notably, characteristic imino-imino cross-peaks from the formation of G:T mismatches are not observed under these experimental conditions, although the chemical shift of the imino proton of G3 is observable (aprox. 11.17 ppm) in the G:T region and exhibits stacking cross-peaks with ^{m}C2 and T22 residues. However, no signal corresponding to exchangeable protons of G16 were found and T1 and T14 could only be sequentially identified. Similar to what we observed for other sequences, we find this fact attributable to a higher flexibility of this structural region provoked by the poly-T linker or to a higher exposure of these residues to the solvent. However, the formation of mixed tetrads is also here supported by the significant chemical shift difference of the amino protons of paired hemiprotonated cytosines. Amino protons of cytosines capped by the G:C base pair of the tetrad are significantly downshifted than those capped by the G:T base pair. Minor groove interaction between guanine residues are unequivocally determined by H1'G8-H1/H22G16 and H1G3-H1'G21 cross-peaks. Stacking contacts between guanine residues and hemiprotonated C:C^{+} pairs are also observed: H5C7-H1/H21/H22G8 and H5/H42C21-H1G21. T4, T9, T17 and T22 methyl protons show non-sequential NOE signals with imino protons from guanine residues, indicating a capping disposition of these residues.

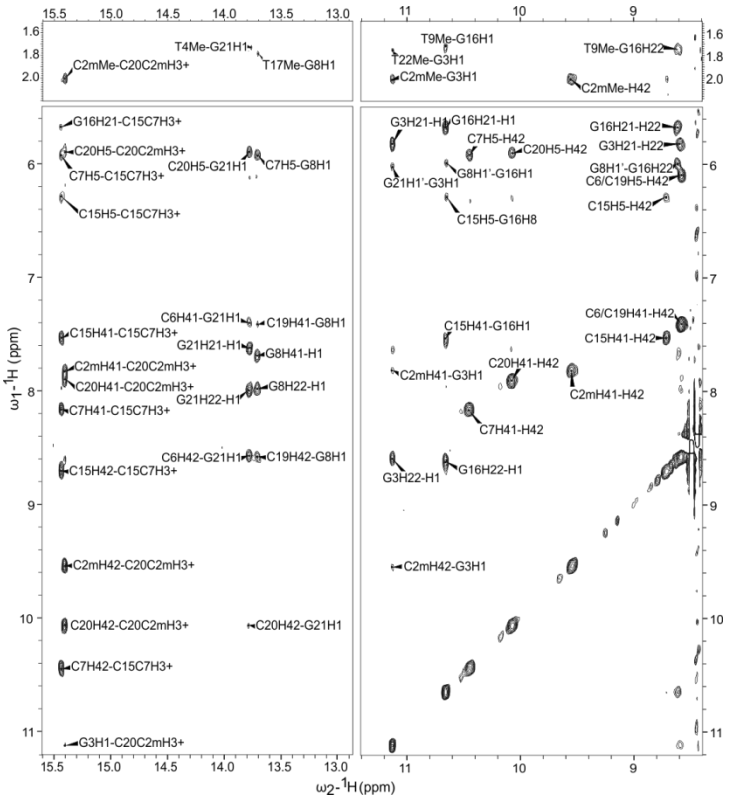


Figure 5-15. Exchangeable protons regions of NOESY spectrum (150 ms) of **LL4_mC(2)** at pH 7 and 5 ºC. H_2O/D_2O 90:10, 10 mM phosphate buffer, [oligonucleotide] = 1 mM.

The full unambiguous assignment of **LL4_^{m}C(2)** was completed on the basis of a large number of sequential cross-peaks in the non-exchangeable proton region. The following contacts match the initially proposed assignment based on the exchangeable protons cross-peaks: H1'T1-Me^{m}C2, H1'/H2"T1-H6^{m}C2, H1'/H2'/H2"^{m}C2-H8G3, H1'G3-MeT4, H8G8- MeT9, H2'/H2"C6/C19-H6C7/C20, H2'/H2"C7-H8G8, H1'C7/C20-H8G8/G21, H2'/H2"C15-H8G16, H2'/H2"C20-H8G21, H1'G16-MeT17, H1'G21-MeT22. The disposition of the hemiprotonated C:C^{+} base pairs at the core of the structure is also well-defined by minor groove interactions reflected in characteristic H1'^{m}C2/C15-H1'C7/20 cross-peaks and by stacking contacts H5C15-H5C20. T5 and T18 residues are degenerated. These residues are exposed to the solvent, but they could be assigned on the basis of sequential sugar contacts: H4'/H5'/H5"T5/T18-H5/H6C6/C19. Remarkably, some contacts involving thymine loop residues are observed: H1'T9-H6T10, H1'T9-MeT10 and Me^{m}C2-H1'Tla. (Tla refers to thymine loop residues that could not be sequentially assigned). All the assigned protons are listed in *Table App* 11.

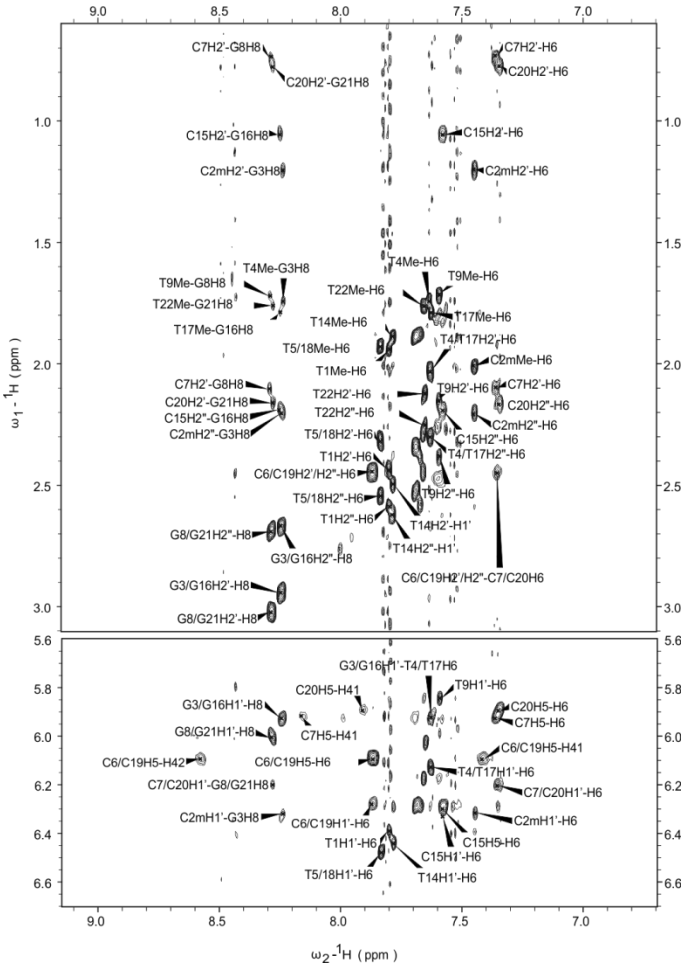


Figure 5-16. Non-exchangeable protons regions of NOESY spectrum (150 ms) of **LL4_mC(2)** at pH 7 and 5 ºC. H_2O/D_2O 90:10, 10 mM phosphate buffer, [oligonucleotide] = 1 mM.

Since **LL4_^{m}C(2)** and **LL4_^{m}C(2,15)** exhibit the same structural features than **LL4**, we can use their assignments to carry out the specific assignment of **LL4**. The comparison of their respective TOCSY spectra, enable the specific assignment of **LL4** by identifying C2 and C15 residues. This approach is graphically shown in *Figure* 5-17, where the H5-H6 regions of **LL4**, **LL4_^{m}C(2)**, **LL4_^{m}C(2,15)** and **LL3** are displayed. From this comparison, it can be easily observed how the cytidine residues assigned to those forming WC G:C base pairs (C6 and C19/C18) present their H6 protons shifted more downfield. Analogously, the H5 protons of C7 and C20 residues (C7 and C19 in the case of **LL3**), which are the ones that were not replaced by ^{m}C residues, are the ones that appear more upshifted.

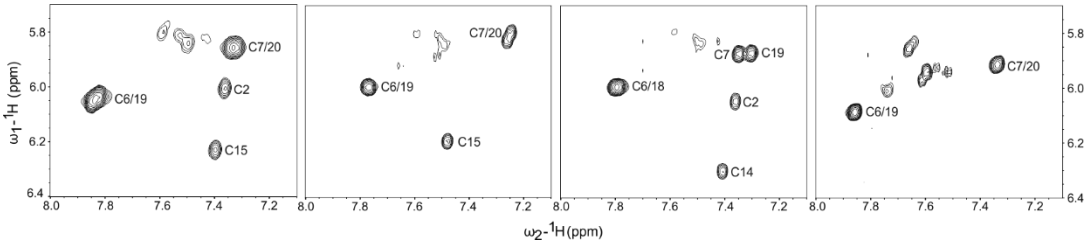


Figure 5-17. H5-H6 cross-peaks region of TOCSY spectra of **LL4** (left), **LL4_mC(2)** (center left), **LL4_mC(2,15)** (center right) and **LL3** (right) at pH 7 and 5 ºC. H_2O/D_2O 90:10, 10 mM phosphate buffer, [oligonucleotide] = 1 mM.

5.3.2.2. **LL3** AND **LL4** ADOPT IDENTICAL HEAD-TO-TAIL MINI i-MOTIF TOPOLOGIES

Based on the unequivocal assignment of the cytidine residues explained on the previous section, the formation of the hemiprotonated base pairs C2:C20$^+$ and C7:C15$^+$ is confirmed in the structure of **LL4** by the contacts between their amino protons and the imino signals shown in the NOESY spectra (15.48 and 15.51 ppm, respectively).

Remarkably, significant differences are found in the chemical shift of amino protons of cytosine residues involved in these C:C$^+$ base pairs. C7 and C20 amino protons are particularly unshielded (10.72 and 10.42 ppm, respectively), showing stacking cross-peaks with the imino protons of G8 and G21 (13.91 and 13.66 ppm, respectively), which, at the same time, show the characteristic cross-peak pattern of WC G:C base pairs with the amino protons of C6 and C19 (G8:C19 and G21:C6). The other two guanine imino signals are found at 12.08 ppm, G16 (exhibiting a broader signal), and at 10.28 ppm, G3. Both signals show cross-peaks with imino protons of two thymine residues: T1 (11.49 ppm) and T14 (11.24), indicating the formation of G:T base pairs (G3:T14 and G16:T1 base pairs). In this case, the amino protons of C2 and C15, which clearly show stacking interactions with these G:T base pairs (H5/H6C-H3T), are shifted upfield compared to those of C7 and C20 (8.82 and 8.60 ppm, respectively). This difference in amino protons chemical shifts of hemiprotonated cytosines stacked on either G:C or G:T base pairs, strongly supports the formation of G:C:G:T minor groove tetrads. Formation of these tetrads is only compatible with a head-to-tail folding. Moreover, formation of G:C:G:T tetrads is also supported by several guanine-guanine cross-peaks: H1'G8-H22G16, H21/H22G8-H22G16. Interestingly, notable different chemical shift are found for the imino protons of G3 and G16.The broad imino signal observed for G16 suggests that this guanine might be more exposed to the solvent than G3, pointing to the possibility that the minor groove tetrads found in these particular motifs are not completely symmetrical, probably due to the effect of the poly-T linker.

Thymine residues T4, T9, T17 and T22 are stacked over the tetrads on the basis of multiple cross-peaks between methyl protons and exchangeable and non-exchangeable protons of guanine residues: MeT4-H8/H1'G3, MeT4-H1G21, MeT9-H8/H1'G8, MeT9-H22G16, MeT17-H1G8, MeT17-H8/H1'G16, MeT22-

H1/H22G3 and MeT22-H8/H1'G21. Moreover, imino signals corresponding to stacked thymine residues are also observed in the 10-12 ppm region: T4 (10.48 ppm), T17 (10.51 ppm), T9 (10.54 ppm) and T22 (10.55 ppm), indicating that these residues are solvent protected. Characteristic i-motif cross-peaks between cytidine residues across the minor groove are also observed: H1'C20-H1'C15 and H1'C6/C19-H5C20/C7.

Sequential assignment based on H6/H8-H2'/H2"/H1' cross-peaks could be completed for T1-C2-G3-T4, C6-C7-G8-T9, T14-C15-G16-T17 and C19-C20-G21-T22 fragments. Thymine residues T1 and T14 can also be identified by their own intraresidual Me-H3 cross-peaks.). Stacked thymine residues (T4, T9, T17 and T22) also show very weak intraresidual Me-H3 cross-peaks. All the assigned protons are listed in *Table App* 12. In *Figure* 5-19, the non-exchangeable protons region at pH 6 is shown. Although the assignment can be analogously followed at pH 7, under slightly acidic pH conditions a better dispersion is observed for thymine aromatic protons of T1 and T14 residues.

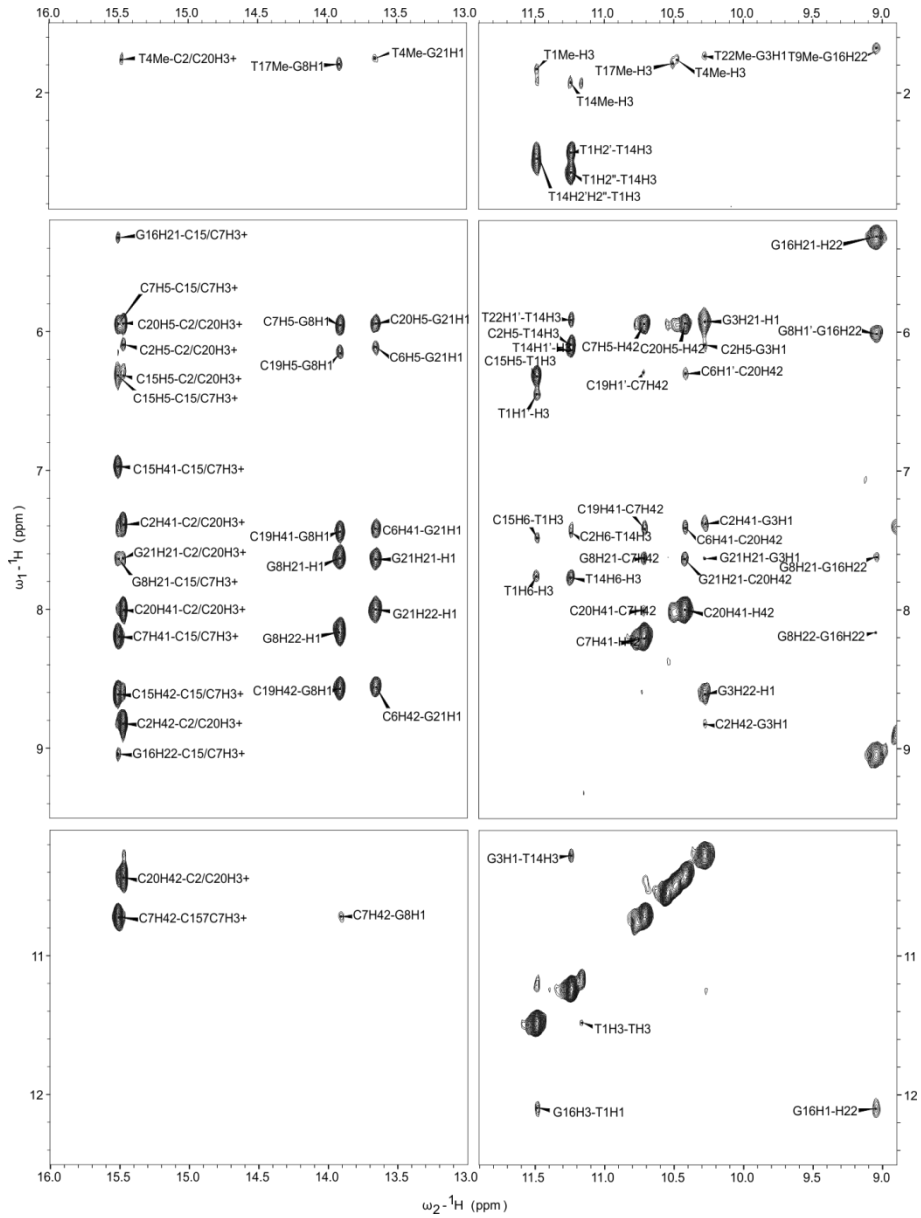


Figure 5-18. Exchangeable protons regions of NOESY spectrum (150 ms) of **LL4** at pH 7 and 5 ºC. H_2O/D_2O 90:10, 10 mM phosphate buffer, [oligonucleotide] = 1 mM.

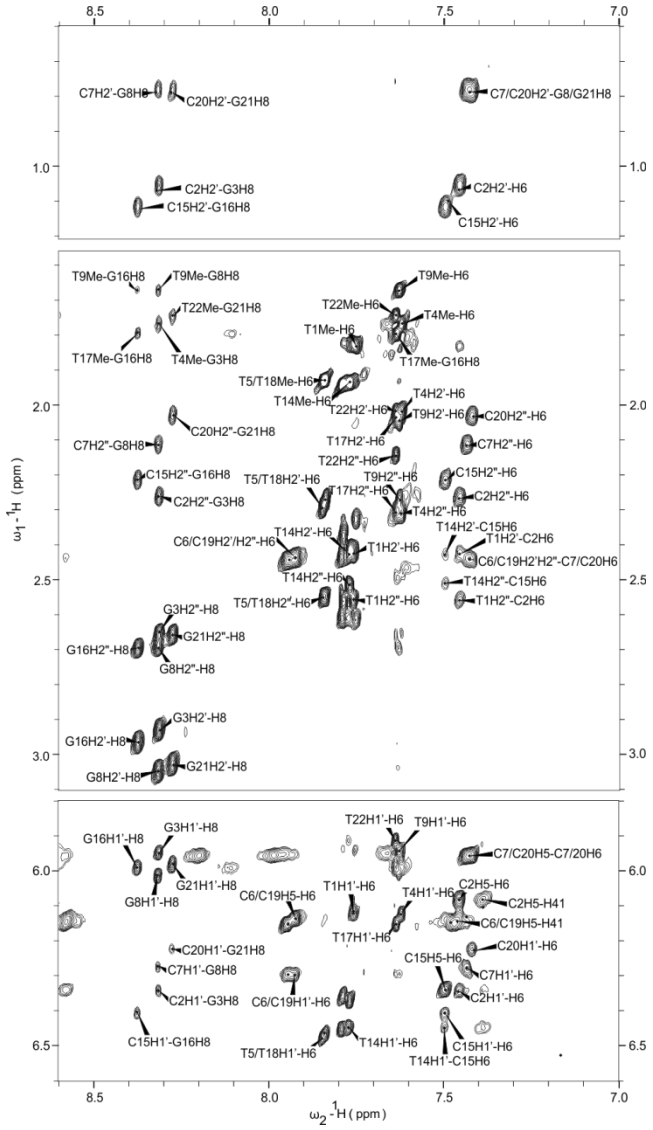


Figure 5-19. Non-exchangeable protons regions of NOESY spectrum (200 ms) of **LL4** at pH 6 and 5 ºC. H_2O/D_2O 90:10, 10 mM phosphate buffer, [oligonucleotide] = 1 mM.

Considering now the NOESY spectra of **LL3** at pH 7, which again are of excellent quality (see *Figure* 5-20), the complete assignment of most of the residues matches the formation of a head-to-tail mini i-motif capped by G:C:G:T tetrads like the previously described structures presenting 5'-TCGTTCCGT-3' repeats. Based on the information presented in *Figure* 5-20, the hemiprotonated base pairs formed in **LL3** i-motif are C2:C19$^+$ and C7:C14$^+$. As in the case of **LL4**, amino protons of cytidine residues from the same hemiprotonated base pair appear with significant different chemical shifts depending on the stacking interactions that they present. The formation of G:C:G:T tetrads is supported by the same contacts described for **LL4**. Two imino protons of guanine residues at 13.94 and 13.65 ppm confirm the formation of WC G8:C18 and G20:C6 base pairs by H1G-H42/H41C contacts. Analogously to **LL4**, that shows a broad signal for the imino proton of G16, the imino proton of G15 is not observed in **LL3**. This information supports the hypothesis of a more exposed guanine residue in one of the tetrads, probably due to a higher flexibility of the 5'-end region of the structure. On the other hand, the G:T mismatch formed between residues G3 and T13 can be perfectly characterized through the H1G3-H3T13 cross-peak. Finally, guanine-guanine contacts through their minor groove side confirm the ensemble of the tetrads (H1'G-H22G).

The sequential assignment of the T1-C2-G3-T4, C6-C7-G8-T9, T13-C14-G15-T16 and C18-C19-G20-T21 fragments of **LL3** could be achieved from the non-exchangeable protons region of the NOESY spectra *Figure* 5-21 following the H6/H8-H2'/H2"/H1' cross-peaks between the residues. Thymidine residues T1 and T13 can be also identified by intraresidual Me/-H3 cross-peaks. Characteristic i-motif cross-peaks between cytidine residues through the minor groove are also observed: H1'C19-H1'C14 and H1'C6-H5C19. In addition, some thymine residues in the loop were assigned on the basis of some Me-H1', but as further confirmation is required, these contacts were not taken into account for structure calculation. Chemical shifts of the **LL3** assignment are listed in *Table App* 13.

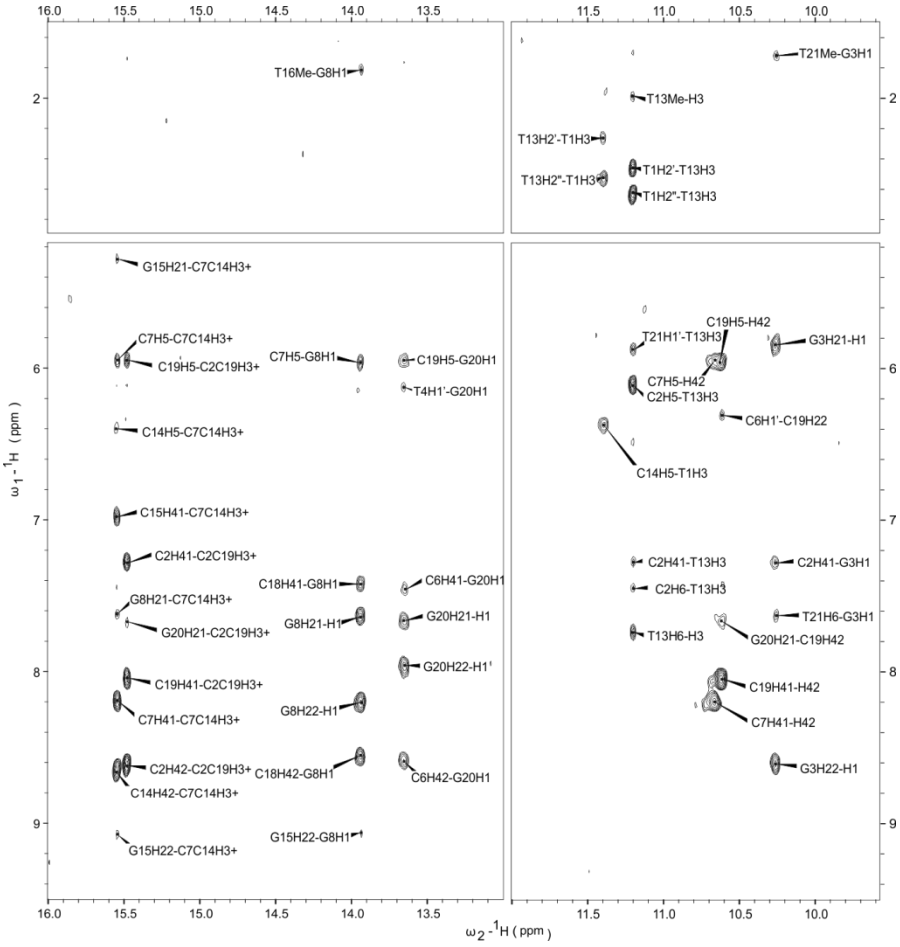


Figure 5-20. Exchangeable protons regions of NOESY spectrum (150 ms) of **LL3** at pH 7 and 5 ºC. H_2O/D_2O 90:10 10 mM phosphate buffer, [oligonucleotide] = 1 mM.

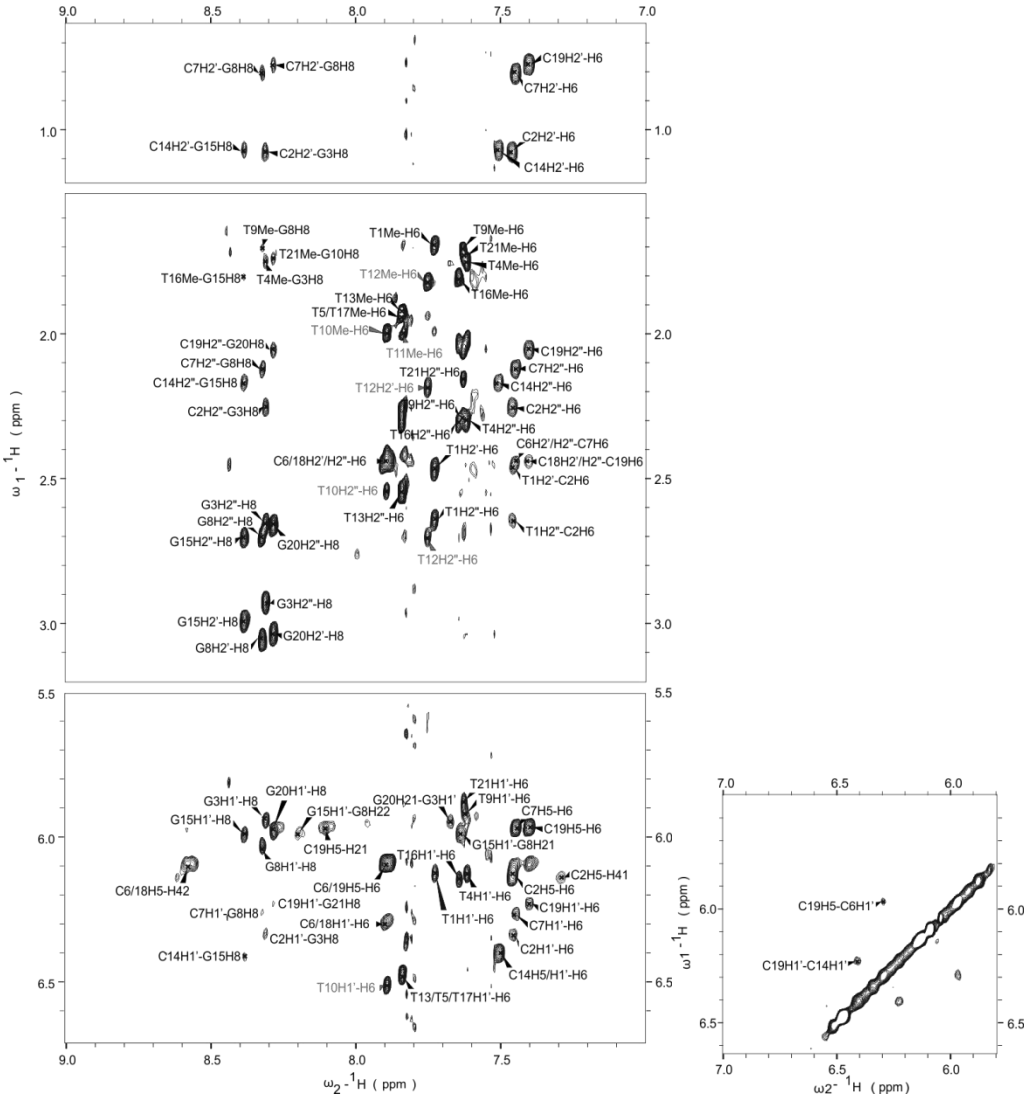


Figure 5-21. Non-exchangeable protons regions of NOESY spectrum (150 ms) of **LL3** at pH 7 and 5 ºC. H_2O/D_2O 90:10, 10 mM phosphate buffer, [oligonucleotide] = 1 mM.

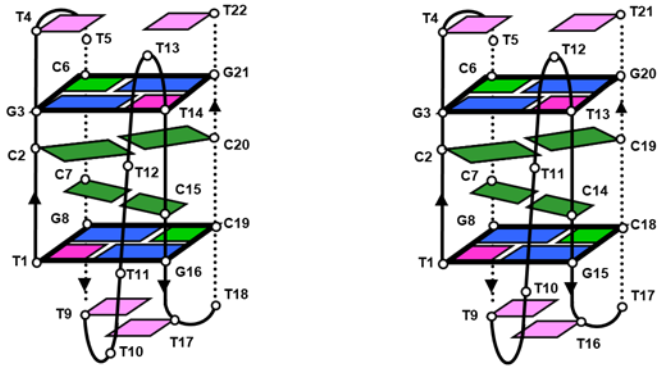


Figure 5-22. Schematic representation of the structures of **LL4** (left) and **LL3** (right) at pH 7.

As a conclusion, the complete assignment of **LL3** and **LL4** supports that both sequences fold into a very similar structure that consists in a mini i-motif structure capped by G:C:G:T tetrads with a head-to-tail topology.

5.3.2.3. SOLUTION STRUCTURE OF **LL3**

The unambiguous assignment of the NOESY spectra for most of the residues of **LL3** and **LL4** at neutral pH confirms that their structures are essentially identical. Therefore, and given the greater thermal stability displayed by **LL3**, we proceeded to determine its three-dimensional structure by restrained molecular dynamics methods.

Based on approximately 125 experimental constraints obtained from the NOESY spectra, the calculations afforded a very well-defined structure with a RMSD value of 0.8 Å. Details on the calculation statistics are reported in *Table* 5-5, and a graphical view of the NOE contacts observed is shown in *Figure* 5-23. Final structures have been deposited at the Protein Data Bank, with PDB code 5OGA. Due do the lack of NOE contacts, certain thymidine residues from the loops and the linker remained poorly defined. The rest of the residues are well-defined in the ensemble of ten structures. The average structure is shown in *Figure* 5-24.

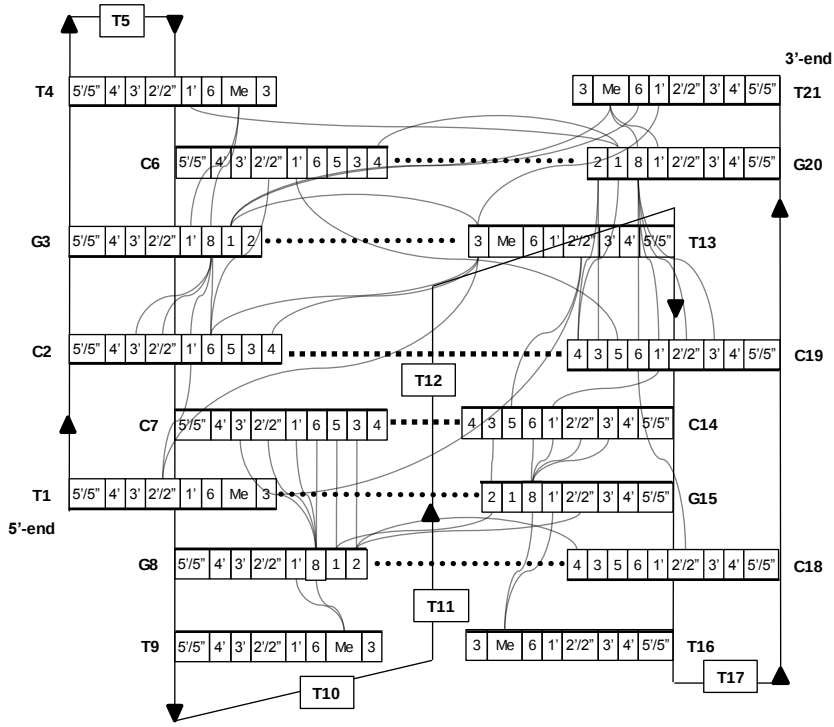


Figure 5-23. Scheme of the most relevant NOE contacts. Round dotted lines indicate G:C and G:T base pairs and square dotted lines stand for C:C$^+$ base pairs.

EXPERIMENTAL DISTANCE CONSTRAINTS		RMSD (Å)		AVERAGE RESIDUAL VIOLATIONS (AND RANGE)	
Total	125	Well-defined bases	0.5 ± 0.1	Sum of violation (Å)	1.23 (0.84-1.69)
Intra-residue	47	Well-defined heavy atoms	0.8 ± 0.2	Max. violation (Å)	0.22 (0.28-0.16)
Sequential	41	Backbone	1.7 ± 0.4	NOE energy (kcal/mol)	5.59 (4.65-6.60)
Range > 1	37	All heavy atoms	2.8 ± 0.7	Total energy (kcal/mol)	-1517 (-1717-1334)

Table 5-5. Experimental constraints and calculation statistics of **LL3**.

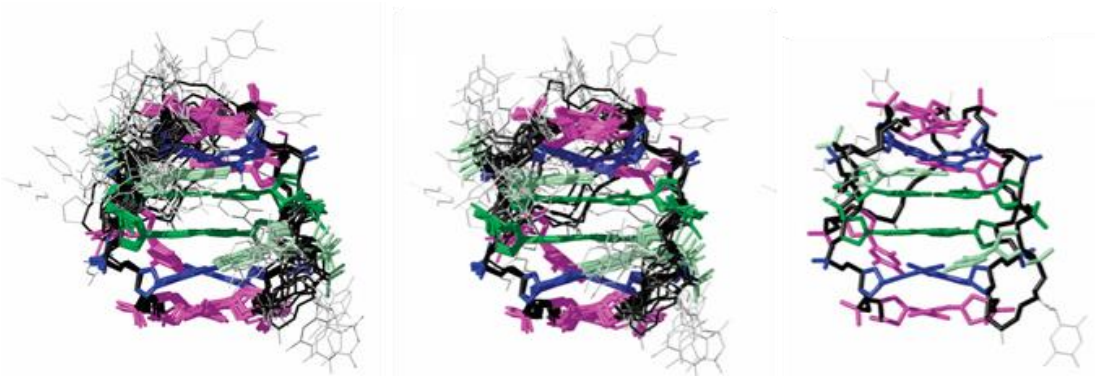


Figure 5-24. Right) Stereoview of the ensemble of ten structures of **LL3**. Left) Average structure. Colour code: Hemiprotonated cytosines are shown in dark green; Watson-Crick cytosines in light green; guanines in blue; well-defined thymines in magenta, and poorly defined residues in gray.

The high-resolution structure of **LL3** confirms all the information obtained from qualitative NMR data in previous sections. The orientation of the oligonucleotidic chain draws a X-shape formed by the two d(TCGTTCCGT) repeats when viewed from the minor groove side (*Figure* 5-25, left). The core of the structure consists of two hemiprotonated C:C$^+$ base pairs capped at both ends by two G:C:G:T slipped minor groove tetrads. The structure is stabilized by a number by hydrogen bond interactions, such as those provided by C:C$^+$, G:C, G:T base pairs and those coming from G:G contacts in the minor grove tetrads. Stacking interactions between the C:C$^+$ base pairs and the guanine residues from the tetrads are also important, as shown in *Figure* 24. This particular interaction is probably key for maintaining the cytidine residues protonated even at high pH conditions. The correlation between stable C:C$^+$ base pairs and the minor groove G:C:G:T tetrads is supported by the fact that the NMR signals from the exchangeable protons corresponding to C:C$^+$, G:C and G:T formation disappear at the same temperature in the NMR melting experiments.

Major and minor grooves in the calculated structures are very peculiar. The formation of minor groove tetrads together with the intrinsic features of i-motif structures, render extremely narrow minor grooves. As it was depictured in the introduction of this thesis, sugar-sugar contacts (which we can observe in the NOESY spectra) between adjacent strands through the minor grooves of the structure, play a significant role for compensating the repulsive interactions generated between nearby sugar-phosphate backbones. Moreover, as we concluded from Chapter 3, the positive charges from hemiprotonated base pairs are also crucial for compensating the electrostatic repulsions between very proximal phosphates.

Regarding the structure of the loops and the poly-T linker, thymidine residues in the first position of the two-reside loops (T4, T9, T16 and T21) are well-defined in the structure and lay on top of their adjacent to guanines. The other thymidine residues from the short loops show not significant NOE cross-peaks and, consequently, are poorly defined in the resulting structures. The residues of the poly-T linker

connecting the two d(TCGTTCCGT) repeats were not included in the restrictions of the calculations since the few cross-peaks that were identified could not be unequivocally assigned.

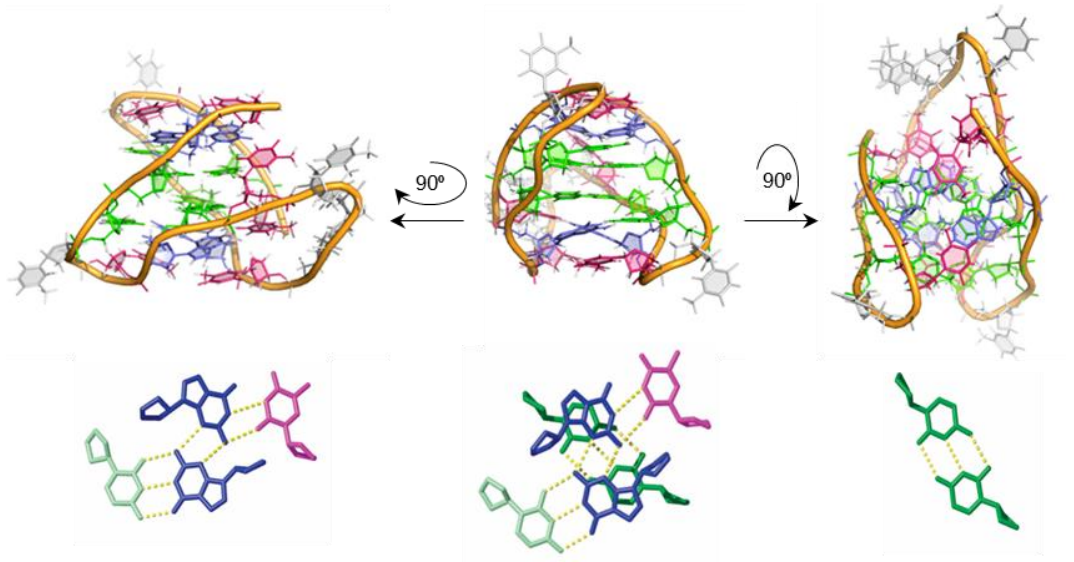


Figure 5-25. Top: Three different views of the calculated structure of **LL3**. Bottom: Structural details discussed in the text. Hemiprotonated cytosines in green, cytosines in WC G:C base pairs in light green, guanines in blue, structured thymines in magenta and exposed thymines in grey.

As reflected by the order parameters shown in *Table* 5-6, the backbone dihedral angles are generally very well-defined, except for some terminal and linker thymine residues. Regarding glycosidic bond, the major configuration exhibited by all residues is anti. However, slight variations throughout the different nucleotides are found. For example, guanine residues present a significant difference compared to the rest of the residues, adopting χ high anti values of around -80º. Also remarkable, are the higher values obtained for the thymine residues stacked over guanines (~ -140º).

The pseudorotation phase angles and the amplitude values that define the sugar puckering in the structure of LL3 are presented in *Table* 5-7. In general, all sugars are in South-East conformations, predominantly in C2'-endo. This differs from the puckering observed for cytosines in other i-motifs, which are generally in C3'-endo or C4'-exo conformations. However, certain residues present a slight deviation to more East configuration adopting O4'-*endo* or C4'-*exo* conformations (i.e. C6 and C18).

Nt	α		β		γ		δ		ε		ζ		χ	
	Avg.	OP	Avg.	OP	Avg.	OP	Avg.	OP	Avg.	OP	Avg.	OP	Avg.	OP
T1	-	-	-	-	58	1.0	87	1.0	173	1.0	-	-	-122	1.0
C2	-55	0.82	-179	1.0	77	0.8	126	1.0	-158	1.0	-86	1.0	-118	1.0
G3	-76	1.0	175	1.0	53	1.0	149	1.0	-115	1.0	-82	1.0	-78	1.0
T4	-83	1.0	176	1.0	52	1.0	148	1.0	-118	0.9	120	1.0	-143	1.0
T5	53	0.34	-179	0.9	66	0.7	143	1.0	-87	0.9	79	1.0	-128	1.0
C6	-151	0.30	-173	1.0	64	0.7	99	1.0	-156	0.7	67	0.4	-119	1.0
C7	-54	0.83	175	0.9	63	0.8	142	1.0	-159	1.0	-123	0.7	-132	1.0
G8	-79	1.0	175	1.0	49	1.0	138	1.0	-103	1.0	-75	1.0	-84	1.0
T9	-93	1.0	-175	1.0	54	1.0	144	1.0	-110	0.8	108	1.0	-144	1.0
T10	-49	0.52	170	0.7	77	0.6	134	1.0	-110	0.8	-149	0.6	-131	0.6
T11	-13	0.20	-178	1.0	63	0.2	137	1.0	-138	0.7	82	0.5	-78	0.2
T12	-64	0.65	169	0.9	62	0.9	128	1.0	-92	0.9	-89	0.6	-126	1.0
T13	68	0.59	169	1.0	66	0.9	136	1.0	-80	1.0	113	0.7	-148	1.0
C14	-85	1.0	172	1.0	48	1.0	140	1.0	-161	1.0	176	1.0	-130	1.0
G15	-74	1.0	173	1.0	58	1.0	149	1.0	-133	1.0	-79	1.0	-82	1.0
T16	-72	1.0	167	1.0	51	1.0	148	1.0	-104	0.9	125	1.0	-145	1.0
T17	91	0.57	178	1.0	50	0.1	146	1.0	-105	0.8	79	1.0	-117	0.9
C18	-171	0.58	-173	1.0	38	0.6	113	1.0	-160	0.9	86	1.0	-117	1.0
C19	-50	0.71	180	1.0	43	0.7	123	1.0	-173	0.8	-131	0.9	-116	1.0
G20	-73	0.93	175	1.0	55	1.0	141	1.0	-126	0.9	-79	0.8	-82	1.0
T21	-83	1.0	175	1.0	54	1.0	145	1.0	-	-	143	0.9	-141	1.0

Table 5-6. Average dihedral angles and order parameters of the structure of **LL3.** Average values correspond to the geometrical mean value and order parameters are use as indicators of angle definition: S=1, perfectly defined and S=0, random distribution.

Nt	Pseudorot.		Sugar Conformation
	Phase	Ampl.	
T1	64	31	C4'-*exo*, O4'-*endo*
C2	137	40	C1'-*exo*, C2'-*endo*
G3	166	36	C2'-*endo*
T4	166	36	C2'-*endo*
T5	153	39	C2'-*endo*
C6	73	29	C4'-*exo*, O4'-*endo*
C7	153	41	C2'-*endo*
G8	150	35	C2'-*endo*
T9	166	33	C2'-*endo*
T10	142	41	C1'-*exo*, C2'-*endo*
T11	147	41	C1'-*exo*, C2'-*endo*
T12	133	39	C1'-*exo*
T13	156	30	C2'-*endo*
C14	169	31	C2'-*endo*
G15	164	36	C2'-*endo*
T16	165	36	C2'-*endo*
T17	157	40	C2'-*endo*
C18	117	33	O4'-*endo*, C1'-*exo*
C19	133	41	C1'-*exo*
G20	165	31	C2'-*endo*
T21	164	34	C2'-*endo*

Table 5-7. Pseudorotation angle and amplitude values that allow the description of the puckering of the ribose ring in the calculated structure of **LL3**.

5.3.2.4. **LL1-7**: THE MINI i-MOTIF STRUCTURE IS COMPATIBLE WITH DIFFERENT LOOP LENGTH

As shown in *Section* 5.2.5, the NMR spectra of **LL1-7** clearly indicate than most structural features in these molecules are independent of the length of the poly-T linker connecting the two repeats, To further confirm this point molecular models of **LL1-7** were built based on the solution structure of **LL3** described in the previous section. These models shown in *Figure* 5-26 were obtained applying the distance constraints obtained for the well-defined residues of **LL3** and varying the length of the unstructured poly-T linker for each structure. Potential linker-structure interactions were not taken under consideration. Although they cannot be taken as an experimentaly determined solution structure, the resulting models show that the main structural features of **LL3** are compatible with loops of different length (**LL1-7**). The most remarkable conclusion obtained from this experiment is that, the globular X-shape disposition of the oligonucleotide backbone allows the formation of this structure independently of the linker connecting the two sequential repeats, even with very short poly-T linkers (**LL1**, **LL2**), which one would not expect to be feasible for a head-to-tail topology.

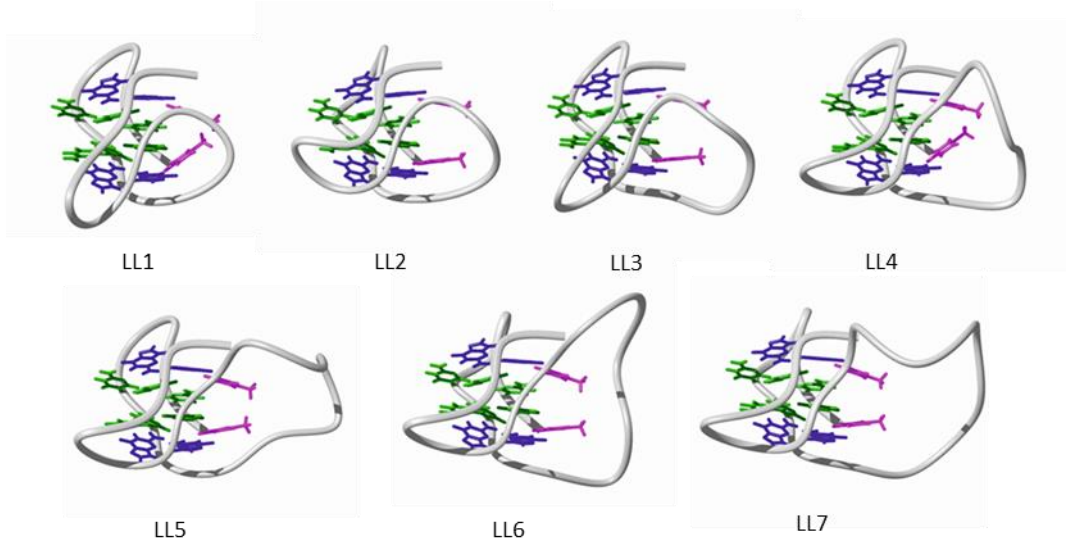


Figure 5-26. Models of the three-dimensional structural models of **LL1-7**.

5.3.2.5. HEAD-TO-TAIL LIKE TOPOLOGIES PREVAIL IN UNIMOLECULAR MOTIFS

In the case of **L**-derived sequences formation of two G:C:G:T or one G:T:G:T and one G:C:G:C minor groove tetrads depends on the topology of the motif. Results obtained from previous and current chapter show that in all cases, the $\mathbf{L_2}$ dimer, its cyclic analogue and **LL1-7** unimolecular mini i-motifs adopt a head-to-tail like orientation that affords G:C:G:T minor groove tetrads formation in all cases over a wide range of pH.

The same behavior is observed for **MM4** and **NN4** sequences that also exclusively fold into head-to-tail structures. Since the nature of these latter sequences provides the same minor groove tetrads independently of the topology of the motif, the high stability of this orientation was attributed to the favorable disposition of the poly-T linker, which lays accommodated on the major groove of the structure. On the contrary, for $\mathbf{M_2}$ and $\mathbf{N_2}$ dimers, an equilibrium between similar populated head-to-tail and head-to-head topologies were found. In the case of LL1-7 sequences, this T-loop-directed head-to-tail topology coincides with the formation of G:C:G:T tetrads, but behavior of $\mathbf{L_2}$ and its cyclic analogue evince that formation of such tetrads is absolutely favored.

This section is focused on evaluating if, in the case of **LL1-7** sequences, the formation of the structure is driven favoring topological factors or, on the other hand, by the nature of the minor groove tetrads. To this end, two different sequences were designed: **LL3HTT1** (d(TCGTTCCGTTTTCCGTTTCGT)) and **LL3HTC1** (d(CCGTTTCGTTTTCGTTCCGT)). These sequences arise from interchanging the C and T residues in positions 1, 6, 13 and 18 from **LL3**. The novelty of these sequences is that, in both cases, the formation of G:C:G:T minor groove tetrads can only occur in a head-to-head orientation of the mini i-motif. Therefore, the study of these sequences will allow to determine whether the head-to-tail topology forces the formation of different minor groove tetrads within the same structure, or the formation of G:C:G:T minor groove tetrads is preferred and, consequently, the i-motif adopts a head-to-head topology.

1D ^{1}H-NMR experiments shown in *Figure* 5-27 confirm the formation of i-motif structures for both sequences at neutral pH. Interestingly, the thermal stability of both sequences is significantly lower than that of the original sequence **LL3**. While the mini i-motif formed by **LL3** shows well-defined i-motif signals at 35 ºC at neutral pH, **LL3HTT1** and **LL3HTC1** only remain folded up to 25ºC. When comparing them, residual signals at 35ºC suggest that **LL3HTC1** is slightly more stable than **LL3HTT1**. Besides the characteristic signals of $C:C^{+}$ hemiprotonted base pairs above 15 ppm, formation of G:C and G:T base pairs is confirmed in both sequences by signals between 13 and 14 ppm and 12 ppm, respectively. Beyond this information, 1D experiments reveal no further hints on the topology of the structures and 2D spectra will be needed to evaluate it.

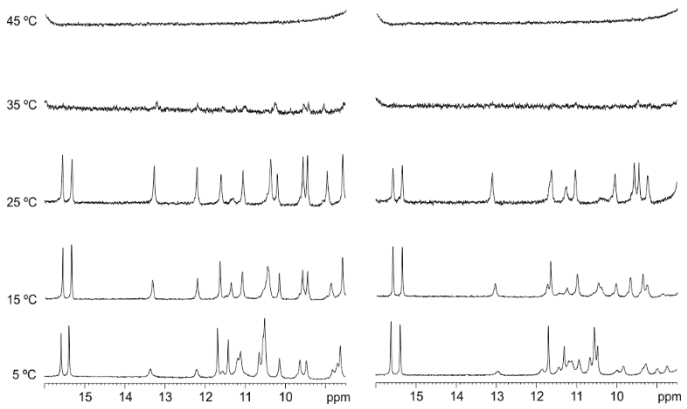


Figure 5-27. 1D ^{1}H-NMR melting series of **LL3HTT1** (left) and **LL3HTC1** (right) at pH 7. 10 mM phosphate buffer (H_2O:D_2O 90:10), [oligonucleotide] = 1 mM.

NOESY spectra recorded at pH 7 and 5 ºC confirm the higher stability of the structure adopted by **LL3HTC1** based on the significant higher quality of the exchangeable protons NOE signals. Consequently, a complete assignment of this sequence has been carried out (see *Figure* 5-28 and *Figure* 5-29).

Six major cytosine H5-H6 cross-peaks are found in the TOCSY spectrum, confirming that one single species is preferred. Four cytosine signals exhibit sequential H8G-H2'/H2" cross-peaks, therefore corresponding to C2, C7, C14 and C19 residues. Two of these residues exhibit sequential cross-peaks with cytosines and were initially assigned to C2/C19 and the other two (C7/C14) to thymine residues. The well differentiated protonated imino signals above 15 ppm corresponded to the formation of the two C:C^{+} base pairs C2:C19^{+} and C7:C14^{+}, consistent to a head-to-tail like folding (see *Figure* 5-28).

Regarding imino protons of guanine residues (*Figure* 5-28), one signal is observed at 13.23 ppm, characteristic chemical shift for G:C WC formation. This signal shows imino-amino cross-peaks with the bonded cytosine and with an hemiprotonated stacked cytosine. A less intense imino signal is observed at 12.04 ppm, too upshifted to belong to a WC bonded guanine but it exhibits quite differentiated amino chemical shifts (8.19 and 5.63 ppm). Other two guanine imino signals ppm, that show characteristic G:T imino-imino cross-peaks with cytosines, are found at 10.97 and 9.99. These two guanines exhibit stacking cross-peaks with C2/C19 cytosines, thus corresponding to G3/G20 residues. These contacts also reinforce the head-to-tail like topology. Formation of minor groove G:T:G:T and G:C:G:C tetrads is supported by amino/imino-H1' contacts between G3 and G21 and, also, between G15 and G8 residues.

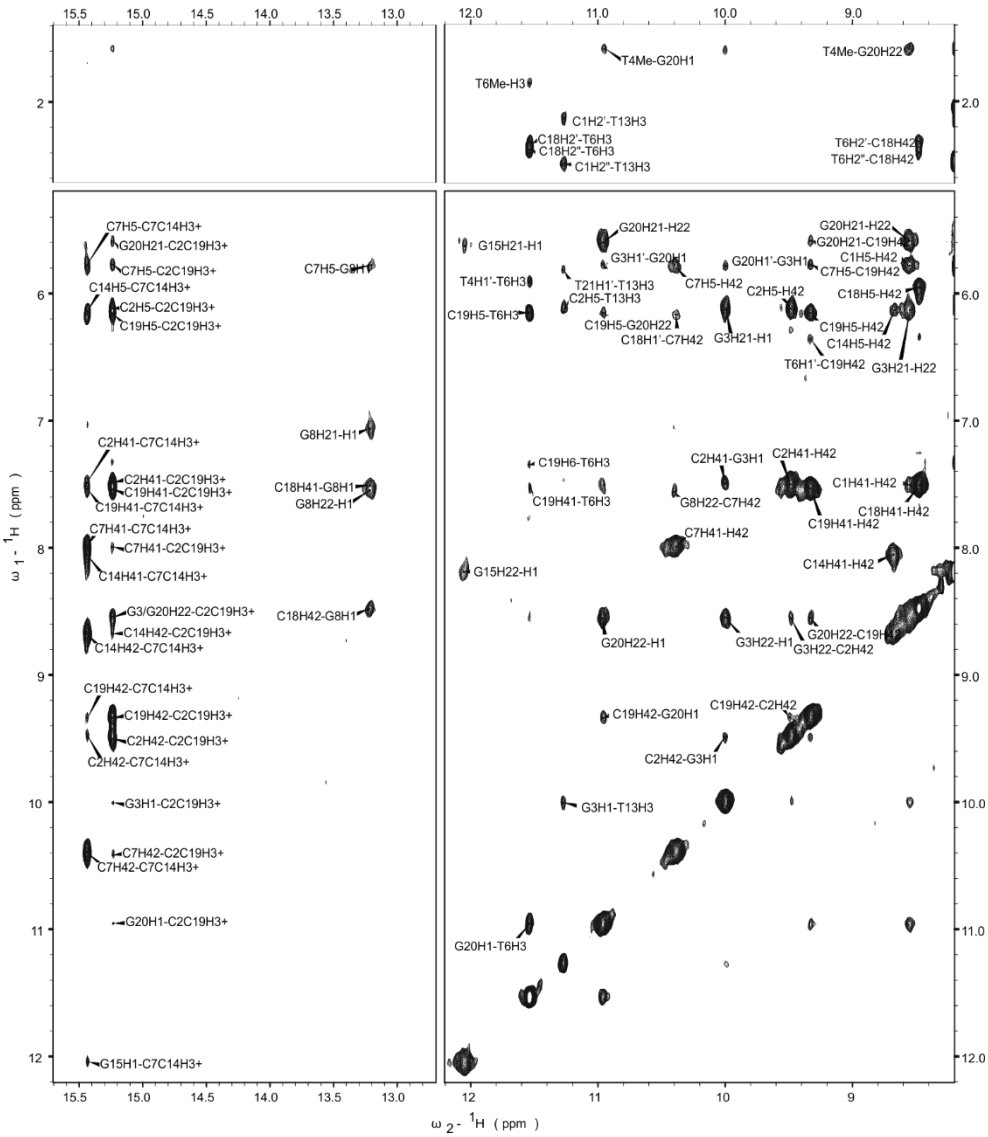


Figure 5-28. Exchangeable protons regions of NOESY spectrum (150 ms) of **LL3HTC1** at pH 7 and 5 ºC. H_2O/D_2O 90:10, 10 mM phosphate buffer, [oligonucleotide] = 1 mM.

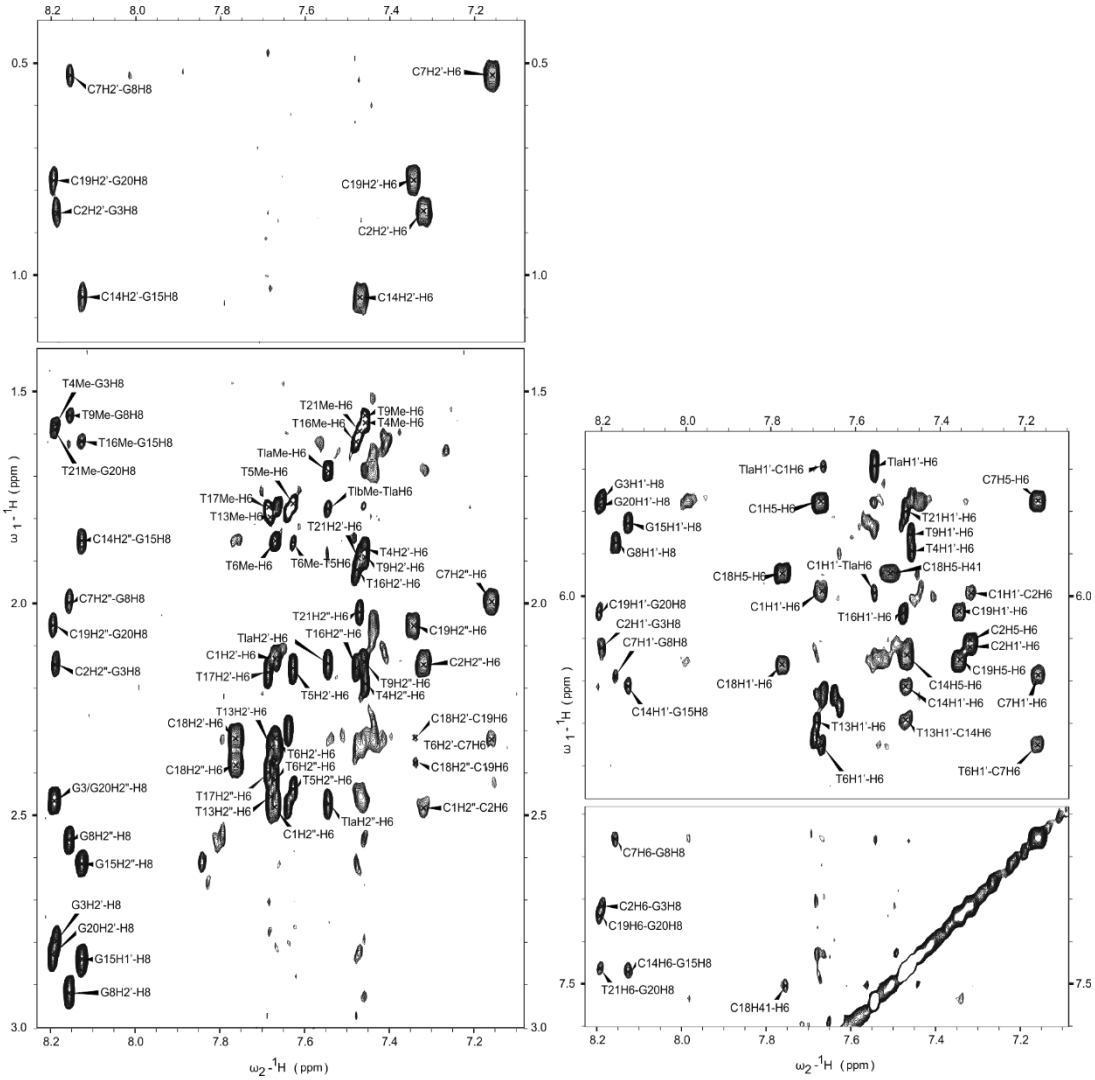


Figure 5-29. Non-exchangeable protons regions of NOESY spectrum (150 ms) of **LL3HTC1** at pH 7 and 5 ºC. H_2O/D_2O 90:10, 10 mM phosphate buffer, [oligonucleotide] = 1 mM.

Specific assignment of the residues could be accomplished on the basis of observed contacts between one of the C1/C18 cytosines with one of the thymines of the connecting loop. According to the head-to-tail structure, the observed MeT-H1'C cross-peak was tentatively assigned to C1. Once identified C1 residue, a complete specific assignment could be carried out, except for the thymines in the loop. The thymine giving contacts with C1 could not be unequivocally assigned to T10 or T11 and it was labelled as Tla (see *Figure* 5-29)

According to this assignment, imino signal at 13.23 ppm corresponded to G8 (G8:C18 base pair), whereas imino proton of G15 exhibited unusual chemical shift for a WC bonding. Although minor groove interaction between G15 and G8 has been observed, it seems like the corresponding G15:C1 WC base pair is not completely formed. Differences between the guanine imino signals involved in G:C base pairs in G:C:G:C tetrads has been also found for N-derived sequences.

Interestingly, for this sequence, contacts between pyrimidine residues in tetrads at both sides could be clearly observed and unambiguously assigned: MeT6-H3'/H2"G8 and Me T13-H2'/H2"C1. These interactions between opposite minor groove tetrads indicate an even more globular shape for mini i-motifs in which tetrads nearly wrap the structure.

In *Figure* 5-30 the schemes of the head-to-tail structures of **LL3HTC1** and **LL3HTT1** are shown. **LL3HTT1** has not been completely assigned but some key cross-peaks pointed to a similar behavior to **LL3HTC1**. A preferred species is also formed at neutral pH as six cytosine H5-H6 cross-peaks were mainly observed. Four cytosines exhibit clear sequential GH8-CH2'/H2" contacts and these cytosines are those involved in $C:C^+$ base pairs formation. On the basis of the sequential contacts, cytosines C2/C19 (previous sequential thymine) were clearly distinguished from C7/C14 (previous sequential cytosine), and hemiprotonated pairs could be assigned to $C2:C19^+$ and $C7:C14^+$ (*Figure* 5-31), confirming the head-to-tail topology. Accordingly, cytosines C7 and C14 showed stacking contacts with guanine residues involved in G:T base pairs. Unfortunately, guanine imino signals in the WC region are only barely observed at these conditions.

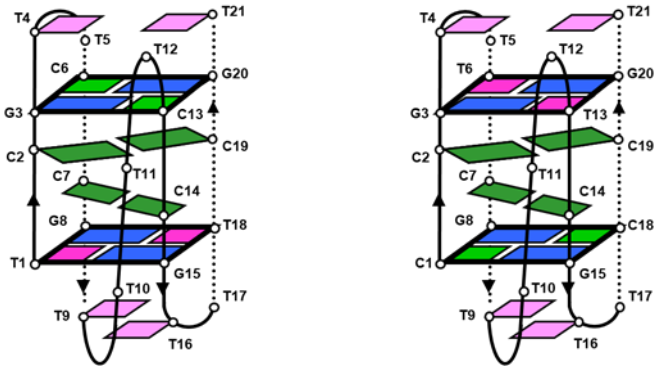


Figure 5-30. Schematic representations of **LL3HTT1** (left) and **LL3HTC1** (right) structures at pH 7.

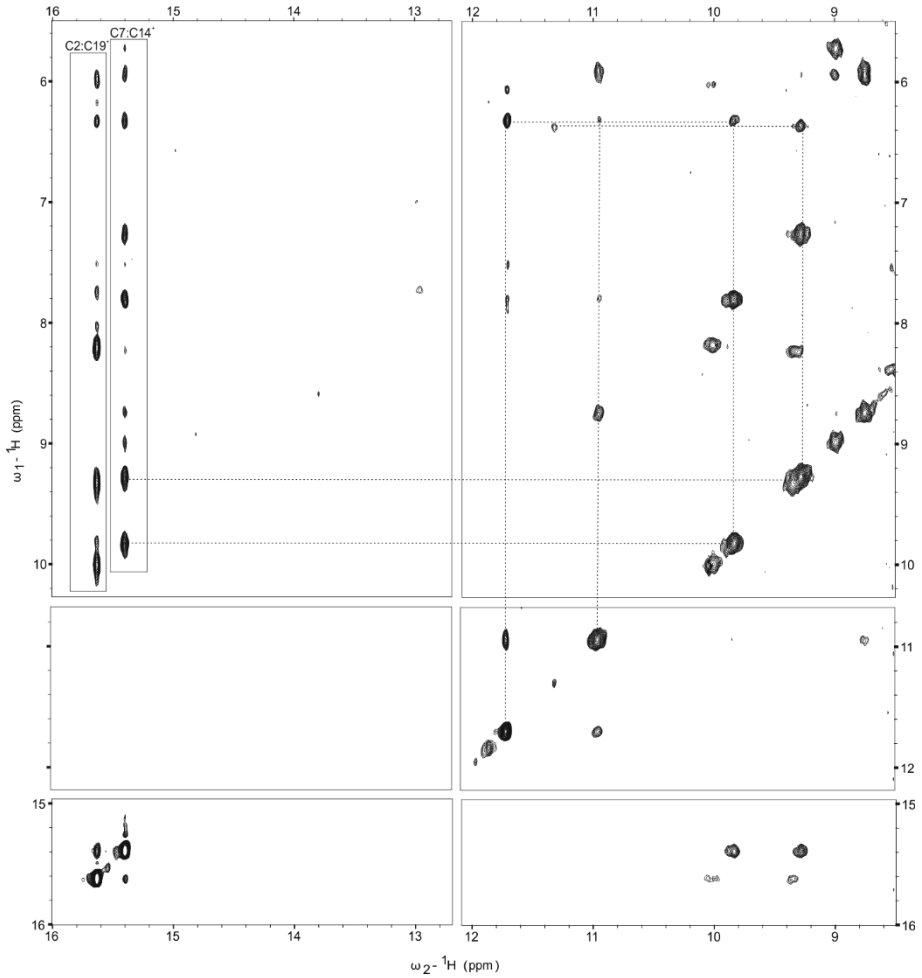


Figure 5-31. Exchangeable protons regions of NOESY spectrum (150 ms) of **LL3HTT1** at pH 7 and 5 ºC. H_2O/D_2O 90:10, 10 mM phosphate buffer, [oligonucleotide] = 1 mM.

5.3.3. NN4 pH-DEPENDENT STRUCTURAL TRANSITION

As mentioned in *Section* 5.2.3, the NMR spectra of **NN4** at neutral pH are not consistent with a single species (*Figure* 5-4). NMR spectra recorded at different pH and temperature indicate the presence of several i-motif species in equilibria. One of these species is favored at higher temperature and pH, whereas the other is predominant at acidic pH. Although, the complete structural determination of i-motifs involved in this equilibrium is part of the doctoral thesis of I. Serrano. I will present in this section preliminary results on the characterization of these species, since these data since the structural insights of **NN4** are crucial for the discussion of the next chapter.

5.3.3.1. **NN4** MINI i-MOTIF IS CAPPED BY G:C:G:C TETRADS AT NEUTRAL pH

As in the case of **MM4**, the highly repetitive sequence of **NN4** provokes an extensive signal overlapping in the NOESY spectra. As shown in *Figure* 5-32, NOESY spectra recorded at pH 7 show the coexistence of alternative i-motif species. One of the two imino signals from hemiprotonated C:C$^+$ base pairs (15.4 ppm) corresponds to two degenerated protons of the major species, whereas the one at 15.7 ppm, which is less intense and shows fewer contacts, belongs to the minor species. Sequential assignments could be done following the same strategy as for **NN4** and **LL4** (i.e. by comparing the original spectra with those resulting from sequence specific C→ ^{m}C mutations). We conclude that the hemiprotonated base pairs formed at this pH are C2:C20$^+$ and C7:C15$^+$, showing clear H41/H42C-H3$^+$C contacts (see *Figure* 5-32). Amino protons of two cytosines (C6 and C19) exhibit cross-peaks with guanine imino protons (13.4 ppm), indicating the formation of two Watson Crick G:C base pairs (G8:C19 and G21:C6). The two signals around 12 ppm were assigned to the imino protons of G3 and G16 based on a number of cross-peaks with the neighbouring C:C$^+$ base pairs (H5C2-H1G3). Although the chemical shifts of these iminio signals are indicative of Watson Crick base pairs, no cross peaks with cytosine amino protons could be detected. Non exchangeable protons of C2 and C15 could be assigned, but their amino protons were not detected. Altogether, the NMR information suggest the formation of two G:C:G:C minor groove tetrads, in which one of the G:C base pairs is strongly distorted. The relative orientation of the two repeats in this structure is head-to-tail, as in the cases of **MM4** and **LL4**. (see *Figure* 5-34).

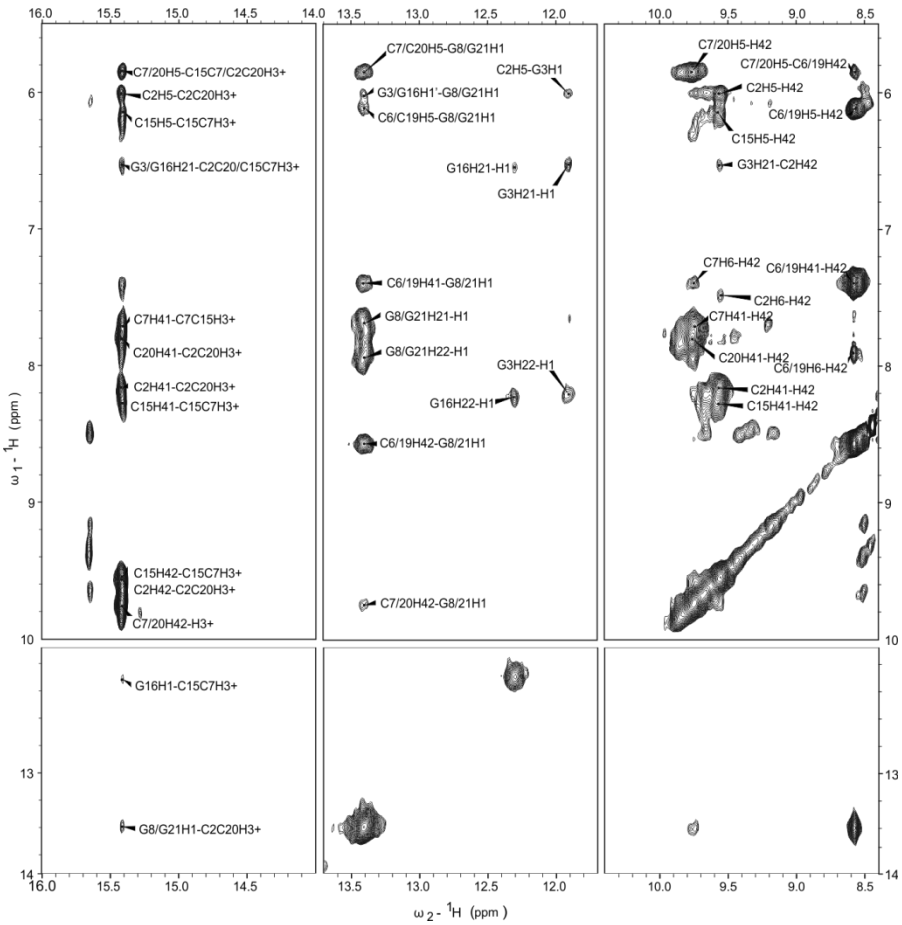


Figure 5-32. Exchangeable protons regions of NOESY spectrum (150 ms) of **NN4** at pH 7 and 5 ºC. H_2O/D_2O 90:10, 10 mM phosphate buffer, [oligonucleotide] =1 mM.

5.3.3.2. **NN4** ACIDIC SPECIES HAS AN EXTENDED C:C$^+$ STACK CAPPED BY G:T:G:T MGT

Spectra of **NN4** at pH 5 render more dispersed signals than those recorded at pH 7, allowing the unambiguous assignment of most of the residues (*Figure* 5-33). The observed signal at 15.7 ppm at pH 7 becomes more intense at pH 5, presenting a high number of cross-peaks. Comparison between the different imino signal intensities suggests the overlapping of two different H3$^+$ signals. Two extra H3$^+$ signals at 15.3 and 15.7 ppm, which also present NOE contacts with two pairs of cytidine amino protons, confirm the formation of four hemiprotonated C:C$^+$ base pairs between non-equivalent residues. The stacking order of these C:C$^+$ base pairs could be determined based on stacking contacts (H5C-H1G), through minor groove contacts (H1'-H1'), and through major groove contacts (H42C1/C14-H2'/H2"C14/C1 and H42C6/C19-H2'/H2"C19/C6). Since no imino signals are observed in the Watson Crick region (12-14 ppm), we can conclude that in this species, C1, C6, C14 and C19 are not involved

in the G:C:G:C tetrads, but forming C1:C19$^+$ and C6:C14$^+$ base pairs. These base pairs, together with C2:C20$^+$ and C7:C15$^+$, form a 4-step stack of hemitprotonated base pairs.

Very interestingly, four imino protons belonging to thymidine residues between 11.0 and 11.5 ppm show NOE cross-peaks with imino protons of guanine residues in the 10.0 to 11.0 region, indicating the formation of four G:T base pairs. Several H1'-H22/H1 contacts between different guanine residues shows the formation of two G:T:G:T minor groove tetrads. The four tymines involved in the two tetrads are T5, T10, T13 and T18. This provokes that the loops are shorter in this species, with two one-residue loops and a two-thymine linker connecting the two subunits. Overall, the NMR data at pH 5 is consistent with the structure shown in *Figure* 5-34.

NN4 at pH 5 adopts a head-to-tail topology with all the structural features previously described. This structure resembles more that of more classical i-motif with a longer C:C$^+$ stack. The increased number of C:C$^+$ base pairs may explain the CD spectra of **N** and **NN4** at acidic pH**,** with maximum bands at around 290 nm. Remarkably, this extension of the C:C$^+$ stack is compatible with a head-to-tail topology in spite of the reduced number of thymidine residues in the linker.

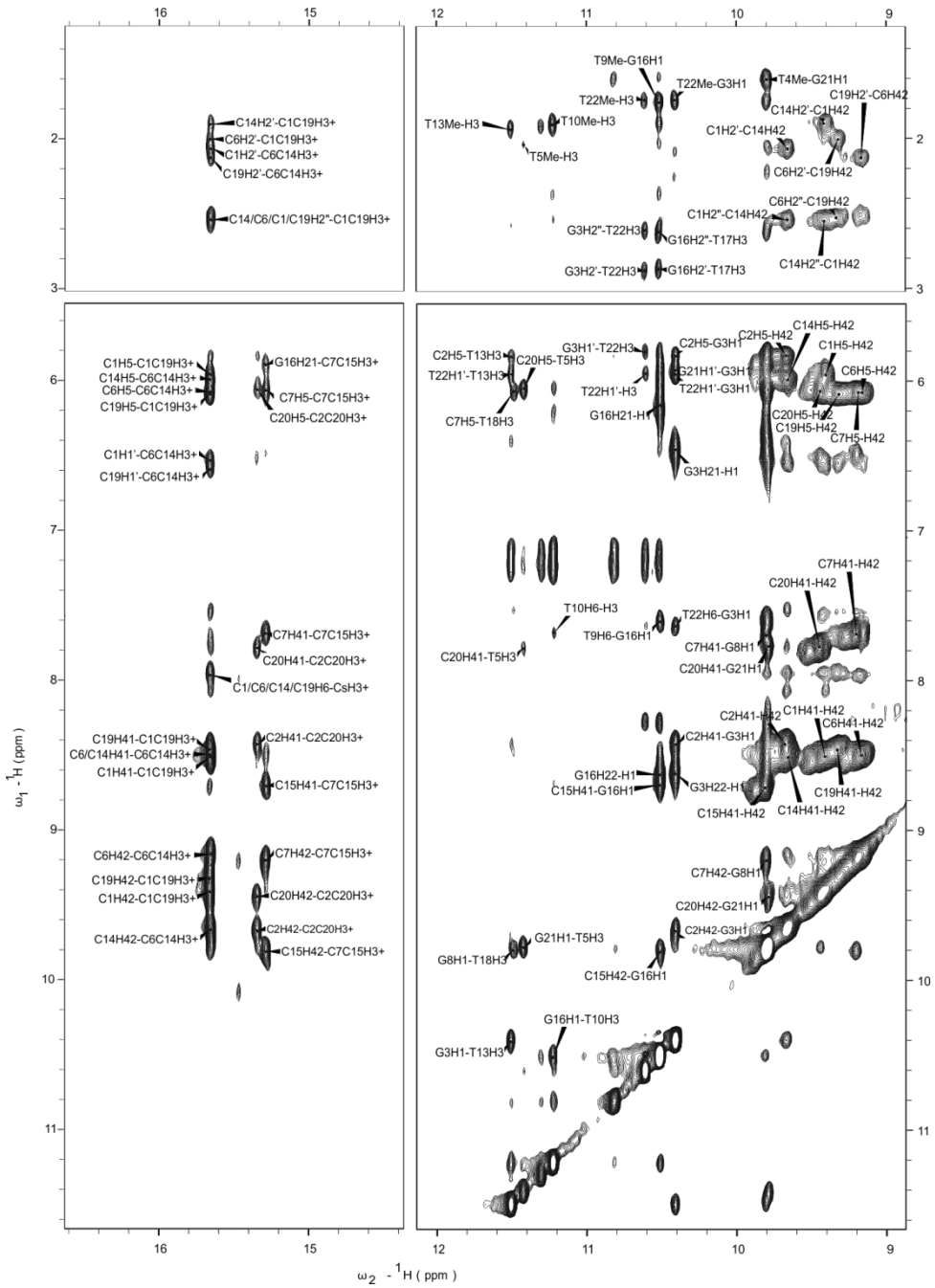


Figure 5-33. Exchangeable protons regions of NOESY spectrum (150 ms) of **NN4** at pH 5 and 5 ºC. H_2O/D_2O 90:10, 10 mM phosphate buffer, [oligonucleotide] =1 mM.

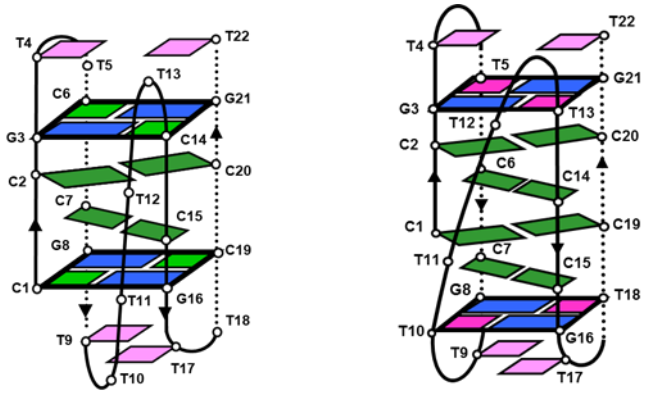


Figure 5-34. Schematic representations of the neutral (left) and acidic (right) species of **NN4**.

5.4. MINI i-MOTIF FORMATION IN TANDEM REPEATS

In order to study the mini i-motif ability to fold into larger tandem constructions, three sequences were used, each containing four repeats of the **L** fragment and different poly-T linkers (see *Table* 5-8). In other words, two longer sequences derived from **LL3** and one longer sequence derived from **LL4** were studied. **LL3rep** and **LL3long** arise from joining two **LL3** fragments with three or six thymidine residues respectively. Analogously, **LL4rep** results from using four thymidine residues to join two **LL4** fragments. The idea behind these constructions is to, as mentioned above, evaluate if the presence of nearby repeats interferes with the formation of mini i-motif structures or, on the contrary, if the structures are solid enough to fold independently in tandem.

As shown in *Figure* 5-35, the previously described longer sequences show very similar spectroscopic properties to their shorter analogues. The exchangeable proton region of the 1D ^{1}H-NMR spectra recorded at pH 7 shows a very well-dispersed set of signals matching the ones observed for the **LL1-7** sequences (see *Figure* 5-8). The formation of hemiprotonated C:C^{+} base pairs is established by several overlapped signals between 15 and 16 ppm. Moreover, imino signals from guanine residues involved in WC base pairing (~14 ppm) and imino signals characteristic to G:T mismatches (11-12 ppm) reveal the formation of G:C:G:T minor groove tetrads. Amino signals from the C:C^{+} base pairs appear between 10 and 11 ppm. This information confirms that the formation of mini i-motif structures in tandem is feasible at neutral pH. Supporting this conclusion, the CD spectra obtained for these longer sequences shows the same characteristic profile obtained for the **LL1-7** sequences with a maximum band between 260 and 270 nm (*Figure* 5-36).

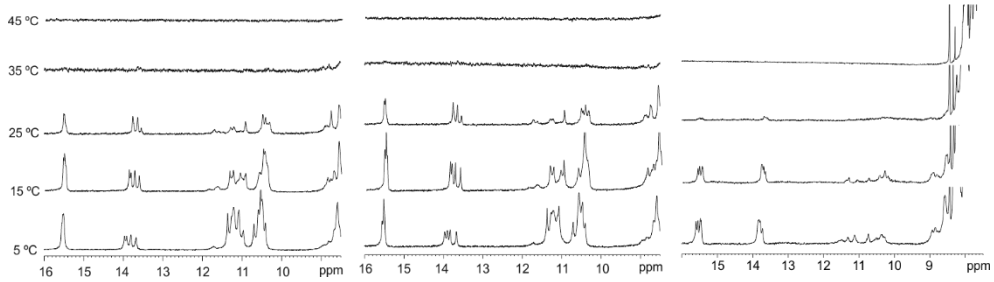


Figure 5-35. NMR melting series of **LL3 rep** (left), **LL3long** (center) and **LL4rep** (right) at pH 7.

In terms of stability, CD and UV spectroscopy experiments afford very remarkable results which remain comparable to the shorter analogues (see *Table* 5-8). Although being slightly lower, the pH_T values calculated from CD-monitored pH titrations for these longer sequences are still quite high for a pH-dependent structure (pH_T values ~ 7.5), reaffirming the strong stabilizing interactions provided by the minor groove tetrads. A more significant decrease in thermal stability is observed from the T_m values obtained from UV melting experiments, which, however, are still above 20 ºC at neutral pH (*Table* 5-8). Entropic factors are obvious to play a more important role in the case of these bigger constructions.

NAME	SEQUENCE	pH_T	T_m @pH7
LL3rep	L-T_3-L-T_3-L-T_3-L	7.6 ± 0.1	23.7 ± 0.5
LL3long	L-T_3-L-T_6-L-T_3-L	7.5 ± 0.1	23.0 ± 0.5
LL4rep	L-T_4-L-T_4-L-T_4-L	7.6 ± 0.1	21.3 ± 0.5

Table 5-8. Sequences of the longer repeats and summary of their stability data (pH_T and T_m values).

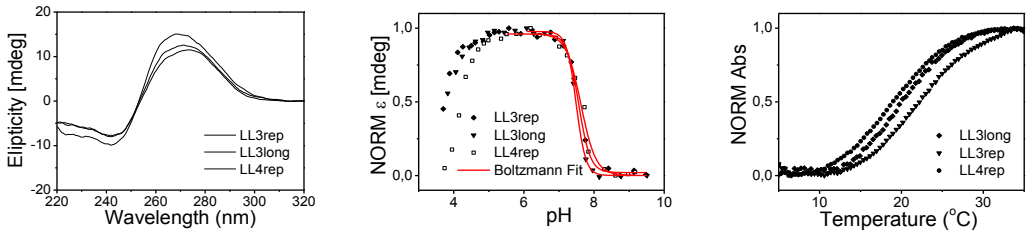


Figure 5-36. CD spectra, titration and melting of **LL3rep**, **LL3long** and **LL4rep**.

Finally, *Figure* 5-37 shows a representation of how we conceive these structures look in tandem, inspired by the solution structure of **LL3**. The hereafter shown model was assembled from an even longer sequence, containing six **L** fragments, forming a mini i-motif structure every two repeats. Due to

the globular shape of the repeating motifs, the tandem formation of these structures resembles bead necklaces which we can imagine to have an important role in their potential biological functions.

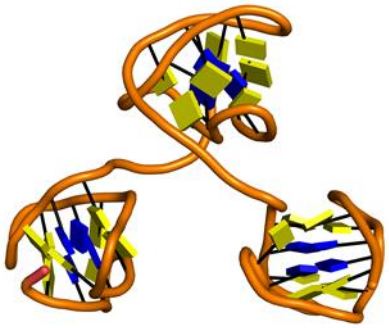


Figure 5-37. Three-dimensional structural model of the formation of three mini i-motifs in tandem.

5.5. BIOINFORMATIC ANALYSIS

Given the unexpected stability of the previously described unimolecular i-motif structures, we wanted to explore if their formation could be feasible *in vivo*. To this end, two different experiments were performed in order to establish the basis on some crucial requirements. First, since these mini i-motif structures would not exist *in vivo* as a discrete unit but as a fragment within a longer, possibly repetitive, DNA sequence, we designed several sequences containing tandem repeats of the 9-mer fragments to compare their structural features and stability. On a different approach, a consensus sequence was mapped to the human genome via bioinformatics analysis in order to monitor its prevalence. Both experiments were performed in collaboration with Modesto Orozco's group and are detailed in the following sections.

5.5.1. CONSENSUS SEQUENCE

From all the conclusions gathered from the stability and structural results presented in this chapter, a consensus sequence was conceived for the formation of unimolecular mini i-motif structures. The most important requirement, common to all the sequences that we studied, is the presence of at least four CG steps. This CG steps are fundamental for, once the structure is folded, afford the high stabilization provided from the interaction between the C:C^+ base pairs and the minor groove tetrads. Interestingly, not only the presence of CG steps is determinant, but the presence of pyrimidine residues at the correct distance is also crucial to allow the formation of the tetrads. Based on previous studies on other structures stabilized by minor groove tetrads,[224] two is the optimal number of nucleotides for the

formation of stable loops connecting the tetrads. Finally, although we observed from the experiments with the **LL1-7** sequences that the length of the linker affects the stability of the motif, its structural features remain unaltered. This information, together with the solution structure of **LL3** that shows that the linker residues remain basically unstructured, we concluded that the nature and the length of the linker are not determinant for the folding of the structure. Altogether, the proposed consensus sequence is the following: d(YCG(XX)YCG(X_n)YCG(XX)YCG), where X can be any nucleotide.

5.5.2. MINI i-MOTIF-FORMING SEQUENCES ARE PREVALENT IN THE HUMAN GENOME

The previously described consensus sequence, d(YCG(XX)YCG(X_n)YCG(XX)YCG), was mapped to the human genome forcing exact matching for n values from 4 to 10 nucleobases. These results were obtained by Diana Buitrago in the group of Prof. M. Orozco. The outcome of this analysis shows 4971 hits for this sequence in the human genome hg19 version (UCSC GRCh37, Feb/2009). Statistically, with a ρ value lower than 10^{-28}, this amount of hits is significantly high compared to a random model. Moreover, as it can be noted in *Figure* 5-38, the number of hits corresponding to sequences not associated to any genes is 118, which is more than five times lower than the expected. Interestingly, the number of nucleotides that resulted more common (statistically twice as expected, ρ value < 10^{-113}) for the length of the linker was 7.

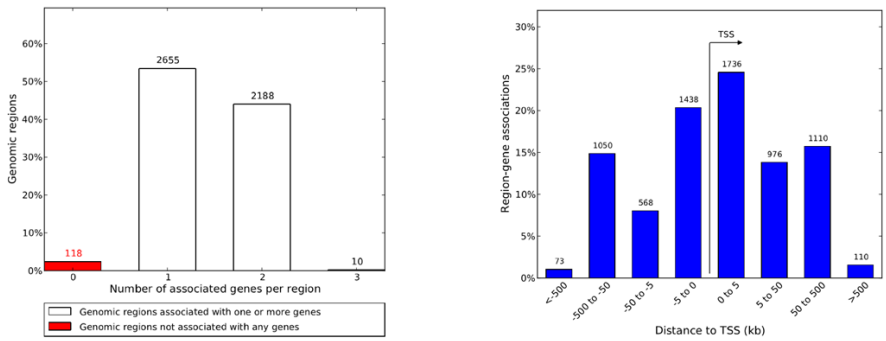


Figure 5-38. Annotation of regions with the consensus motif using Genomic Regions Enrichment Annotations (GREAT) Tool. Left: Number of hits associated to 0, 1, 2 or 3 genes, taking minimum distance upstream (5kb) and downstream (1kb) of the TSS. Right: Distribution of distances between hits and TSS of their associated genes.

Remarkably, *Figure* 5-38 *(right)* shows that mini i-motif-forming sequences appear to be not randomly distributed in the human genome, but very significantly localized in regulatory regions, gathered very close to transcription starting sites (TSS). More specifically, these sequences are found over-represented in promoter regions and 5'UTRs and poorly represented in introns and intergenic regions (see *Table* 5-9*)*. This information reveals that, as it was described in the introduction of this thesis for

well-known i-motif-forming sequences, where the equilibrium between B-DNA and i-motif structures regulates gene expression, a possible function of mini i-motif structures could involve the regulation of transcriptional processes. Furthermore, this hypothesis is also supported by a gene ontology (GO) analysis,[225] that shows that the proposed consensus sequence is found over-represented, with great statistical confidence, in genes with development functions (see *Table* 5-10).

ANNOTATION	HITS	TOTAL SIZE (bp)	Log2 ENRICHMENT
3'UTRs	28	23005177	-0.384
ncRNA	17	6409543	0.74
pseudo	11	2010139	1.785
Exon	358	36768443	2.616
Intron	1239	1240886382	-0.669
Intergenic	841	1749354789	-1.724
Promoter	2129	35414190	5.243
5'UTRs	264	2760547	5.912

Table 5-9. Genomic annotation of hits. Enrichment of hits for each category is computed as the ratio between the normalized number of hits (number of hits in a category divided by the total number of regions) and the normalized number of base-pairs per region type (percentage of base-pairs of the genome per category). Genomic sequences shorter than 1 Mbp were excluded to reduce noise.

TERM NAME	FDR p-VALUE	FOLD ENRICHMENT	OBSERVED REGION HITS
Regulation of metanephric glomerulus development	7.94876e-23	10.4771	38
Positive regulation of kidney development	2.36364e-6	2.8127	39
Enteric nervous system development	2.77212e-6	2.5710	45
Rac protein signal transduction	8.82346e-5	2.8911	28
Phospholipid translocation	1.16460e-4	2.6258	32
Intrinsic apoptotic signaling pathway in response to endoplasmic reticulum stress	2.04422e-3	2.7949	21
Paraxial mesoderm morphogenesis	2.65032e-3	2.8223	20
Corpus callosum development	4.54575e-3	2.8760	18
Paraxial mesoderm development	2.48624e-2	2.1246	24
Negative regulation of protein kinase activity by regulation of protein phosphorylation	3.17901e-2	4.7062	7

Table 5-10. GO Enrichment analysis. Functional annotation terms significantly enriched at hits of the consensus motif as discovered by GREAT. For each term we report the p-value corrected for multiple testing (FDR p-value), the ratio of observed/expected regions with the annotation (Fold Enrichment) and the number of regions mapped to genes associated to the GO term.

5.6. CONCLUSIONS

The designed oligonucleotide sequences **MM4**, **LL4** and **NN4** all fold into unimolecular mini i-motif structures at neutral pH stabilized by G:T:G:T, G:C:G:T and G:C:G:C minor groove tetrads. The topology of all the structures was proven to be predominantly head-to-tail, indicating a preference of the poly-T linker to rest accommodated through one of the major grooves of the structure.

The solution structure of these oligonucleotides at pH 7 was precisely obtained based on experimental data, showing a highly compact construction with a very characteristic globular X-shaped topology.

In terms of stability, **LL4** and **NN4** render extraordinary results, affording stable i-motif structures even at basic pH. On the other hand, **MM4** thermal stability at neutral pH is rather low. This information allows to rank the stabilizing effect provided by different minor groove tetrads as it follows: G:C:G:T ≥ G:C:G:C >> G:T:G:T.

Longer sequences containing a higher number of repeats show significantly similar results compared to their shorter analogues, confirming that the formation of these motifs in tandem is feasible and that their stability is remarkably high.

Bioinformatics analysis show an overexpression of the consensus sequence designed to adopt mini i-motif structures throughout the human genome. Specifically, this study revealed that the localization of these sequences is not randomly localized, but gathered close to TSS. Moreover, the genes in which these sequences are found, majorly have regulatory and development functions.

Some of results and conclusions obtained from this piece of work were recognized as worth to be published.[226]

6. FLUORESCENT MINI i-MOTIFS

6.1. BACKGROUND AND OBJECTIVES

6.1.1. STRUCTURAL APPLICATIONS OF FLUORESCENCE IN NUCLEIC ACIDS

Fluorescent tools are especially relevant for the study of non-canonical nucleic acid structures since they allow, for instance, the characterization of intra- versus intermolecular folding, the study of the dynamic equilibria between different species or the description of the kinetics of structural transition processes. The previously described studies are not always accessible by other spectroscopic techniques and thus, fluorescence must be considered as a very important technique for the structural study of nucleic acids. Some of the strategies for optimizing the applications offered by this technique include the use of molecular beacons,[166] FRET systems,[165] fluorescent nucleobases[164,227] and other fluorescent probes.[167,228,229]

Regarding the use of fluorescent nucleobases (some of them shown in *Figure* 6-1), the use of 2-aminopurine (2AP) was first reported in 1969[230] and continues being the most used fluorescent analogue in the context of DNA characterization. 2AP advantages are based on the high quenching observed in the fluorescence emission upon DNA hybridization. This characteristic behavior of 2AP has been remarkably used, for instance, for the characterization of DNA dynamics at local level[231] or for the detection of base flipping events in DNA-enzyme complexes.[232,233] Other well-known fluorescent nucleobases include the family of pteridine analogues 3MI, 6MI, 6MAP or DMAP, for which the quenching observed depends on the proximity and number of purine residues within the oligonucleotide strand.[234,235] Other examples include the use of pyrene derivatives that significantly enlarges the range of wavelengths used for the excitation of the fluorescent system.[236–238]

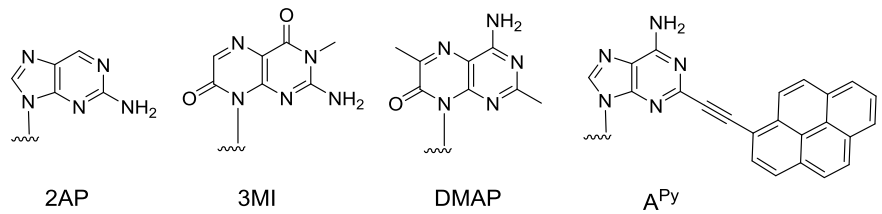


Figure 6-1.Some of the most well-known purine-derivative fluorescent nucleobases.

In the case of i-motif structures, a great number of nanotechnological and biomedical applications have been described by using fluorescence spectroscopy taking advantage on their pH-dependent nature,[239] some of which were already described in the introduction of this thesis. In terms of structure characterization and dynamics, a considerable number of studies report the use of FRET systems[240–242] and time-resolved fluorescence applications[243,244] in i-motif constructions. However, nucleobase analogues that are able to hybridize as a cytosine residue are of special interest since they can mimic the formation of C:C$^+$ base pairs, and thus may provide very specific and deeper insights on the

dynamics of the formation of i-motif structures. Some of the examples that have been proposed include a quinazoline derivative (^{DMA}C) used for the kinetic study of the transition between DNA duplexes and i-motif structures[228] or phenoxazine derivatives (tC, tC^o) used for labeling i-motif structures within structural transitions[243,245,246] or as pH fluorescent probes.[247] In fact, the previously described i-clamp moiety, used as a nucleobase modification for the stabilization of i-motif structures, is a direct derivative of tC^o.[100] *Figure* 6-2 shows the chemical structure of these fluorescent analogues.

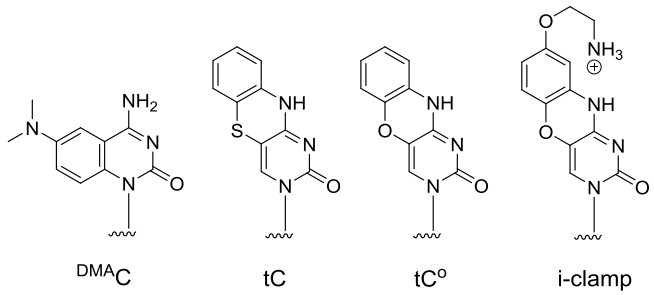


Figure 6-2. Best-known fluorescent cytidine analogues used for the study of i-motif structures.

6.1.1.1. TRICYCLIC 1,3-DIAZA-2-OXOPHENOXAZINE DERIVATIVE: tC^o

Among the different fluorescent cytosine analogues previously used for the study of i-motifs, tC^o exhibits a number of advantages: Compared to its structurally equivalent analogue tC, tC^o shows a better quantum yield with absorption maximum at 370 nm (see *Figure* 6-3).[248] Moreover, it has been reported that both phenoxazine derivatives do not significantly affect the secondary structure of B-DNA upon hybridization.[248,249] The high brightness displayed by tC^o upon hybridization in G:tC^o WC-like base pairs formation, affords remarkably low quenching of the fluorescence signal in a B-DNA duplex environment.[248] On the contrary, when incorporated into an i-motif-forming sequence, the formation of tC^o:C$^+$ base pairs provokes a high quenching of the fluorescence signal displayed by this fluorophore.[243] Very interestingly, both tC and tC^o can be strategically used in FRET systems as donors combined with the FRET-acceptor analogue tC$_{nitro}$.[241,250] Finally, the commercially available phosphoramidite form of tC^o allows its easy incorporation to an oligonucleotide chain by standard SPOS methodology (see 7.2.4 in Methods).

Consequently, we considered that incorporation of tC^o in our mini i-motif structures would be helpful for a deeper structural analysis.

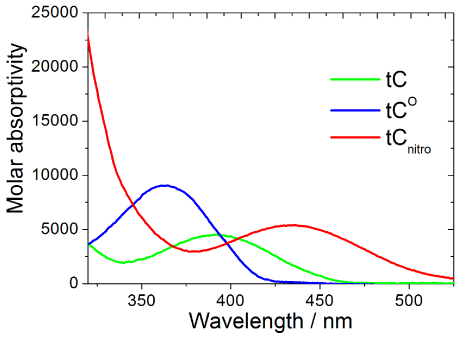


Figure 6-3. Molar absorptivity of tC, tC° and tC_{nitro} as a wavelength function. Source: Glen Research.

6.1.2. SEQUENCES

The idea of incorporating a fluorescent analogue of cytosine in a mini i-motif structure was conceived in the context of the equilibrium between different i-motif species found for the sequence **NN4**. As it was described in the previous chapter, the unimolecular mini i-motif capped by G:C:G:C minor groove tetrads described for **NN4** at neutral pH, exhibits an equilibrium with an alternative i-motif structure with an extended $C:C^+$ core and G:T:G:T minor groove tetrads that becomes prevalent at pH 5. Interestingly, in NN4 mini i-motif, cytosine residues in different structural an chemical environment are found and protonation state of some them changes upon interconversion between the two structures On this basis, we thought of **NN4** as a very interesting candidate for testing the feasibility of fluorescently label a mini i-motif structure.

More in detail, the studies proposed in this chapter would ideally lead us to determine if:

1- Given the great stability exhibited by **NN4**, the substitution of cytosine residues differently hybridized within the mini i-motif structure (WC hydrogen bonded or involved in hemiprotonated base pairs) by tC° units is compatible with the formation of the motif.

2- By means of fluorescence studies, the incorporation of tC° allows the detailed characterization of the structural transition observed for **NN4** acting as an internal structural probe.

To this end, two **NN4**-derived sequences were designed (see *Table* 6-1). **NN4_tC°(2)** incorporates a tC° moiety in the position 2 of the sequence **NN4**. Based on the solution structure of **NN4** at pH 7 (*Figure* 6-4, left), if the formation of the mini i-motif structure remains unaltered upon the incorporation of tC°, the structure of **NN4_tC°(2)** should include a $tC°2:C7^+$ hemiprotonated base pair (*Figure* 6-4, center). On the other hand, **NN4_tC°(6)** incorporates a tC° residue in the sixth position of the **NN4** sequence. This substitution, if not destabilizing, should ideally result in the formation of a WC-like G21:tC°6 base

pair involved in the formation of a minor groove G:C:G:C tetrad (*Figure* 6-4, right). *Figure* 6-5 shows the chemical structure of expected tC°:C$^+$ base pair and G:tC°:G:C minor groove tetrad.

NAME	SEQUENCE
NN4	d(CCGTTCCGT-TTTT-CCGTTCCGTT)
NN4_tC°(2)	d(CtC°GTTCCGT-TTTT-CCGTTCCGTT)
NN4_tC°(6)	d(CCGTTtC°CGT-TTTT-CCGTTCCGTT)

Table 6-1. Sequences studied in this chapter.

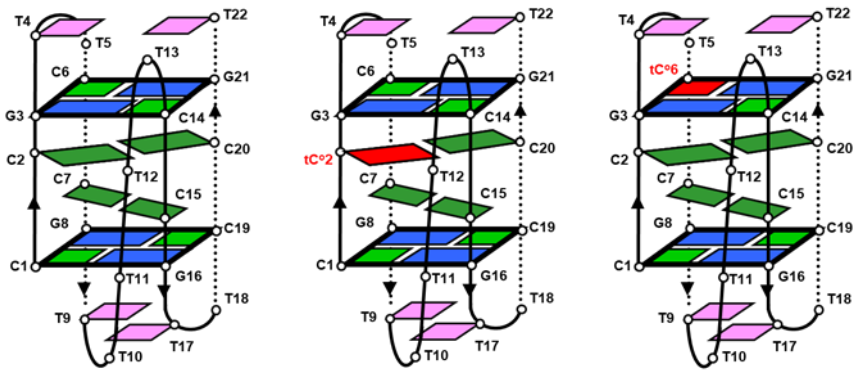


Figure 6-4. Schematic representation of the expected structures adopted at neutral pH by the sequences studied in this chapter. **NN4** (left), **NN4_tC°(2)** (center) and **NN4_tC°(6)** (right).

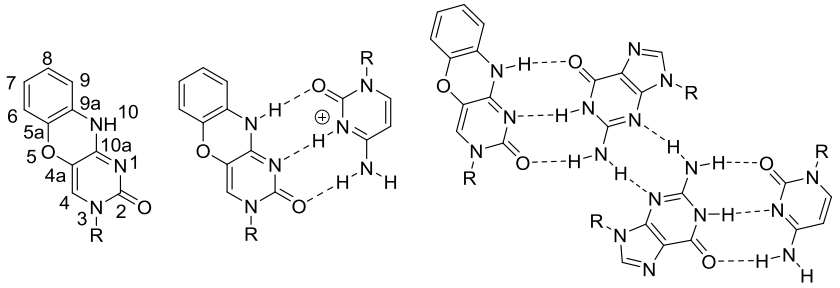


Figure 6-5. Numeration of the tC° residue (left), tC°:C$^+$ base pair (center) and G:tC°:G:C minor groove tetrad (right).

The oligonucleotides utilized in this chapter were synthetized by standard SPOS. The results of the syntheses are summarized in *Section* 7.2.4 in Syntheses.

6.2. FORMATION AND STABILITY

NMR, CD, UV and fluorescence techniques will be used in order to assess and discuss the formation and thermal and pH-stability of i-motif structures containing the fluorescent nucleobase tC^o. Previously reported results for the parent **NN4** sequence are also shown for comparison.

6.2.1. 1D ^{1}H-NMR STUDIES AT pH 7

As can be observed in *Figure* 6-6, 1D ^{1}H-NMR melting series at pH 7 reveal the formation of i-motif structures for both **NN4_tC^o(2)** and **NN4_tC^o(6)** sequences. Comparable signal dispersion and pattern in the exchangeable proton region to that of the spectra of **NN4** is observed, suggesting the formation of a very similar structure. For **NN4_tC^o(2)**, an extra intense signal in the hemiprotonated resdues region is observed, probably corresponding to the tC^o:C$^+$ base pair. Analogously to **NN4**, characteristic signals from WC G:C base pair between 13-14 ppm are observed for **NN4_tC^o(2)**. Also, signals at 12-13 ppm indicate the presence of hybridized guanine residues as in **N** and **NN4** that might not match canonical WC pairing. Imino signals from guanine residues are also observed at unusually upshifted chemical shifts. In the amino region (10-12 ppm), a greater number of signals in the tC^o-containing sequences spectra are observed compared to those of **NN4**, possibly including the H10 proton of the tC^o residue.

Notably, signals corresponding to the formation of i-motif structures for the sequence **NN4_tC^o(2)** are observed at higher temperatures values (up to 45 ºC) than those of **NN4** and **NN4_tC^o(6)**, pointing towards a stability enhancement in this sequence provoked for the specific incorporation of tC^o.

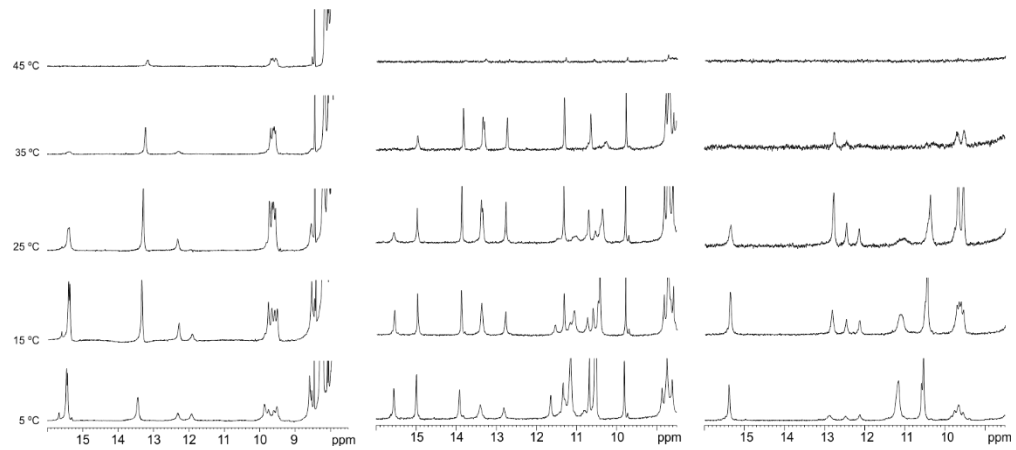


Figure 6-6. 1D ^{1}H-NMR melting series of **NN4**, **NN4_tC^o(2)** and **NN4_tC^o(6)** at pH 7. 10 mM phosphate buffer (H_2O:D_2O 90:10), [oligonucleotide] = 1 mM.

6.2.2. A STRUCTURAL TRANSITION IS OBSERVED UPON ACIDIFICATION

NMR spectra recorded at acidic pH, shown in *Figure* 6-7, confirm that the designed tC^o-containing sequences **NN4_tC^o(2)** and **NN4_tC^o(6)** follow the same behavior of **NN4** and fold into an alternative i-motif structure at acidic pH values. As in the non-modified sequence, minor signals around 15 ppm become moderately intense at lower pH values. The observation of these extra imino signals occurs under stronger acidic conditions for **NN4_tC^o(2)** than for the other **N**-derived sequences. As for the other **N**-derived sequences, the spectra under acidic conditions reveal that no WC G:C base pairs are formed and that increasing number of signals around 11 ppm possibly correspond to the formation of G:T mismatches

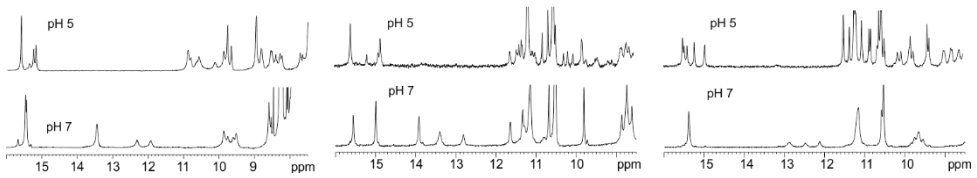


Figure 6-7. NMR spectra of **NN4**, **NN4_tC^o(2)** and **NN4_tC^o(6)** at different pH values (5 and 7) and 5 ºC. 10 mM phosphate buffer (H_2O:D_2O 90:10), [oligonucleotide] = 1 mM.

In order to provide an initial characterization of the different neutral and acidic structures adopted by these sequences, CD spectra were recorded at pH 5 and pH 7 (see *Figure* 6-8).

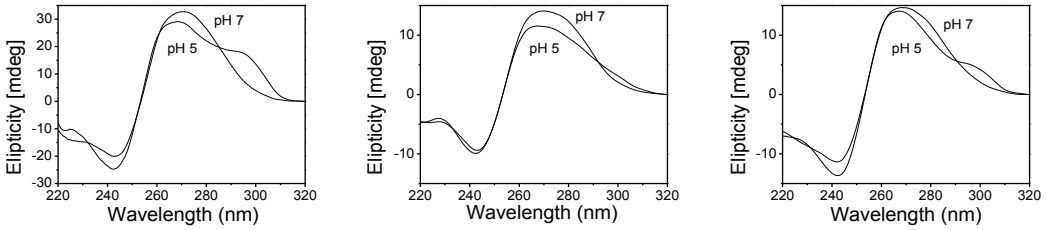


Figure 6-8. CD spectra at different pH values of of **NN4**, (left) **NN4_tC^o(2)** (center) and **NN4_tC^o(6)** (right) at different pH values (5 and 7). 25 mM phosphate buffer, [oligonucleotide] = 2.0 µM.

Analogously to what was observed for **NN4**, a significant difference is observed in the CD spectra of the tC^o-containing sequences recorded at acidic and neutral pH. In both conditions, the spectra share the same wavelength for the maximum (~ 270 nm) and minimum (~ 240 nm) ellipticity values. However, the spectra recorded at pH 5 are of slightly less intensity and, most remarkably, exhibit a characteristic i-motif band near 300 nm. This information, as in the case of **NN4**, is consistent with the formation of a more classic i-motif-like structure for **NN4_tC^o(2)** and **NN4_tC^o(6)** sequences at acidic pH. Noticeably, the intensity of such band at 300 nm in the spectra of **NN4_tC^o(2)** at pH 5 is lower compared to that of

the other sequences, suggesting a difference in the dynamics of the conformational transition between the neutral and acidic species respect to **NN4** and **NN4_tC^o(6)**. These structural transitions will be studied in detail by fluorescence spectroscopy in *Section* 6.4.

6.2.3. tC^o POTENTIALLY ENHANCES THE THERMAL STABILITY OF i-MOTIF STRUCTURES

As it was suggested by the NMR melting experiments shown in *Figure* 6-6, the incorporation of tC^o in mini i-motif-forming sequences confers great thermal stability to the resultant i-motif structures. Interestingly, a different degree of stabilization was observed depending on the position in which the fluorescent analogue is incorporated, reminding of the results obtained for the protonated cytidine analogue psC.

In order to quantify the potential of tC^o as a stabilizing agent of i-motif structures, UV melting curves were recorded at neutral and acidic pH values. The results obtained from these experiments are summarized in *Table* 6-2. **NN4** T_m values are included for comparison.

NAME	SEQUENCE	T_m pH 5	T_m pH 7
NN4	d(CCGTTCCGT-TTTT-CCGTTCCGTT)	53.3 ± 0.5	29.2 ± 0.5
NN4_tC^o(2)	d(CtC^{o}GTTCCGT-TTTT-CCGTTCCGTT)	57.7 ± 0.5	38.8 ± 0.5
NN4_tC^o(6)	d(CCGTTtC^{o}CGT-TTTT-CCGTTCCGTT)	53.4 ± 0.5	30.7 ± 0.5

Table 6-2. Summary of the stability results (T_m values were calculated from UV melting experiments shown in *Figure* 9) obtained for **NN4**, **NN4_tC^o(2)** and **NN4_tC^o(6)**. 25 mM phosphate buffer, [oligonucleotide] = 2 μM.

Very interestingly, as can be observed in the melting profiles shown in *Figure* 6-9, a notable positive difference is observed for the sequence **NN4_tC^o(2)**. After adjusting the experimental data to a sigmoidal Boltzmann fit, the calculated T_m values reveal a significant enhancement of the thermal stability of the i-motif structure adopted by **NN4_tC^o(2)** at pH 7 of approximately +10ºC. On the other hand, the thermal stability displayed by **NN4_tC^o(6)** is very similar and comparable to that of **NN4**. Overall, and as expected for i-motif structures, the T_m values of all the sequences notably increase (+20ºC, approx.) at pH 5. Under these experimental conditions, the same tendency observed at pH 7 regarding thermal stability is observed. However, in this case, the gap between **NN4_tC^o(2)** and **NN4** and **NN4_tC^o(6)** is lower (<5ºC). Interestingly, at acidic pH, although T_m values for **NN4_tC^o(6)** and **NN4** are quite similar, melting profiles exhibit different shape, showing **NN4_tCo(6)** a higher cooperativity.

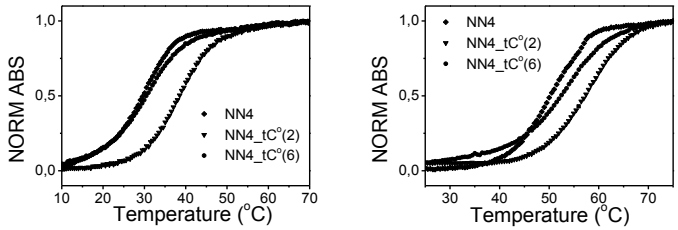


Figure 6-9. UV melting curves obtained for **NN4**, **NN4_tC°(2)** and **NN4_tC°(6)** at pH 7 (left) and pH 5 (right). 25 mM phosphate buffer, [oligonucleotide] = 2.0 μM.

In order to test the efficiency and reliability of fluorescence applied to the study of this particular sequences, excitation and emission spectra were recorded, as it is explained in *Section* 8.7.2 at different temperature values for **NN4_tC°(2)** and **NN4_tC°(6)** at pH 7 (*Figure* 6-10/11) and pH 5 (*Figure* 6-12/13). As expected for tC°-containing sequences, maximum absorption bands are found at 370 nm and maximum emission bands at 450 nm.

Remarkably, from the experiments performed at pH 7, one can observe that at 5 ºC, conditions under which the i-motif structure should be folded for both sequences, while the fluorescence signal is heavily quenched for the sequence **NN4_tC°(2)**, for the sequence **NN4_tC°(6)** is practically unquenched. This information is consistent with our initial hypothesis of tC° forming a tC°:C$^+$ base pair in **NN4_tC°(2)** (*Figure* 6-5, center), quenched due to hemiprotonation, and a WC-like G:tC° base pair in **NN4_tC°(6)** (*Figure* 6-5, right), whose fluorescence does not exhibit significant changes between folded and unfolded species. Remarkably, the recovery of the fluorescence signal due to the unfolding of the structure of **NN4_tC°(2)** takes place at approximately 35 ºC, which matches the stability data obtained by UV spectroscopy.

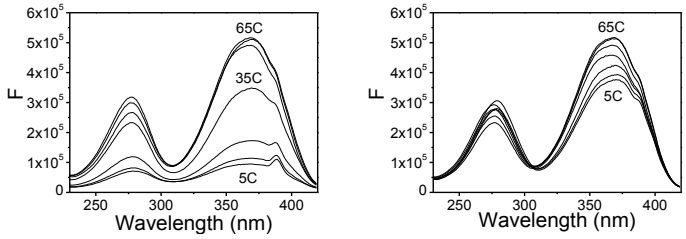


Figure 6-10. Fluorescence excitation spectra recorded at different temperature for **NN4_tC°(2)** (left) and **NN4_tC°(6)** (right) at pH 7. 25 mM phosphate buffer, [oligonucleotide] = 0.2 μM.

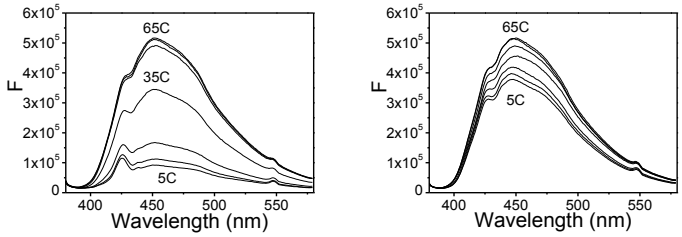


Figure 6-11. Fluorescence emission spectra recorded at different temperature for **NN4_tC°(2)** (left) and **NN4_tC°(6)** (right) at pH 7. 25 mM phosphate buffer, [oligonucleotide] = 0.2 μM.

Analogously to what is observed at pH 7, fluorescence is differently quenched for **NN4_tC°(2)** and **NN4_tC°(6)** at pH 5. For **NN4_tC°(2)**, similar quenching profile to that obtained at pH 7 is observed but, as expected, the recovery of fluorescence occurs at higher temperature values (55ºC approx.). In the case of **NN4_tC°(6)**, in comparison to pH 7, a higher quenching is observed at low temperature values , suggesting a different chemical environment of tC° residue under these experimental conditions These results for **NN4_tC°(6)** are consistent with the reported **NN4** i-motif structure at acidic pH in which protonation of cytosines in the tetrad occurs leading to additional C:C$^+$ base pairs.

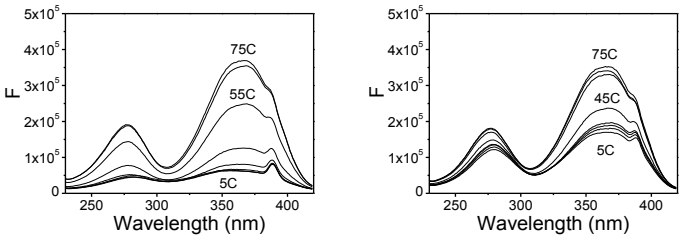


Figure 6-12. Fluorescence excitation spectra recorded at different temperature for **NN4_tC°(2)** (left) and **NN4_tC°(6)** (right) at pH 5. 25 mM phosphate buffer, [oligonucleotide] = 0.2 μM.

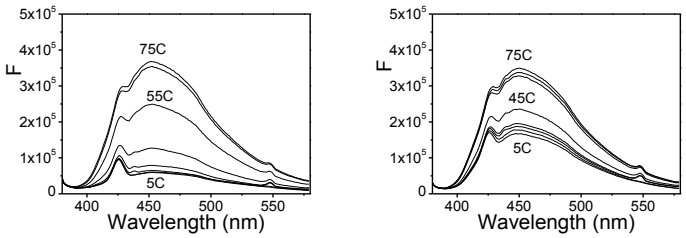


Figure 6-13. Fluorescence emission spectra recorded at different temperature for **NN4_tC°(2)** (left) and **NN4_tC°(6)** (right) at pH 5. 25 mM phosphate buffer, [oligonucleotide] = 0.2 μM.

6.3. STRUCTURAL CHARACTERIZATION

The results presented in the previous section provide some interesting hints that suggest that **NN4_tC^o(2)** and **NN4_tC^o(6)** i-motif structures are of the same nature than the ones described in the pH-driven structural equilibrium for **NN4**. In order to confirm this hypothesis and to complete the full characterization of these structures, 2D NMR experiments were recorded at neutral and acidic pH values. Moreover, since the ultimate objective of this chapter is to detail the structural transition between different i-motif species by means of fluorescence spectroscopy, a special attention will be given to the characterization of the diverse hybridization states in which the tC^o residue is found within the different structures.

6.3.1. NN4_tC^o(2)

6.3.1.1. tC^o IS COMPATIBLE WITH THE FORMATION OF MINI i-MOTIF STRUCTURES

The increased thermal stability observed for **NN4_tCo(2)** at neutral pH renders high quality NMR spectra with a great number of NOE cross-peaks and a good signal dispersion. The expected number of cytosine and thymine residues for a unimolecular species is confirmed from the signals observed in the TOCSY spectrum which were unequivocally identified as ten Me-H6 thymine contacts and five H5-H6 cytosine protons cross-peaks. The aromatic system of tC^{o}2 residue can be also easily distinguished in the TOCSY spectrum.

In the exchangeable proton region of the NOESY spectra, two protonated imino signals are observable at 15.57 and 15.00 ppm. The signal at 15.00 ppm was assigned to a tC^{o}2:C$^+$ base pairs, as it shows cross-peaks with a couple of cytosine amino protons and the H10 proton of tC^{o}2, whereas the other signal corresponds to an unmodified C:C$^+$ base pair (see *Figure* 6-14).

In the case of **NN4**, site-specific substitutions of cytosine by 5-^{m}C were used to determine the folding topology of the motif. In this case, tC^o contacts were essential for determining the folding pattern of the structure. The H10 proton of tC^{o}2 shows cross-peaks with a couple of cytosine H2'/H2" protons, which are characteristic contacts between 3'-3' (back-to-back) intercalated cytosine residues in i-motif structures. Since this cytosine is not hemiprotonated, these signals should correspond to that WC-hydrogen bonded to G3. Both, this cytosine and tC^{o}2 show cross-peaks with the same thymine residue (tC^{o}2H8-H1'/H2'/H2"/H6T, tC^{o}2H7-H2'/H2"T, TH2'/H2"-CH5 and TH1'/H2'/H2"-H5/H6C) thus, it necessarily has to be a thymine residue from the loop probably located in the major groove region. All these contacts are consistent with a head-to-tail folding of the i-motif in which mentioned residues would correspond to C14 and to an undetermined T residue from the central linker. The complete assignment of the sequence is consistent with this topology.

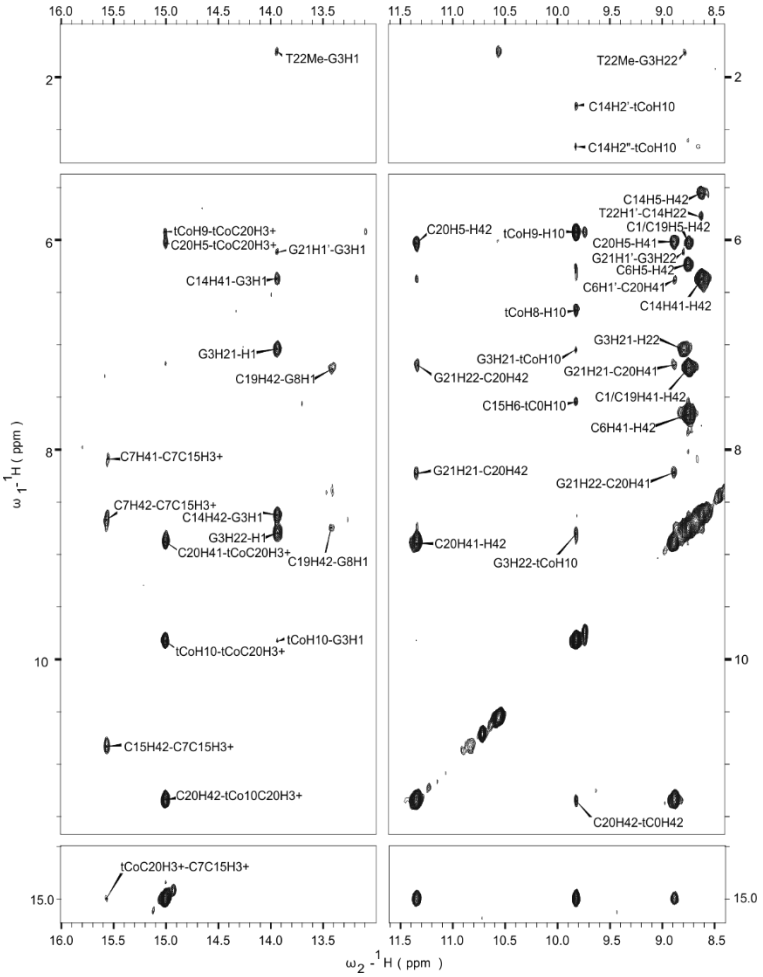


Figure 6-14. Exchangeable protons regions of NOESY spectrum (150 ms) of **NN4_tC^o(2)** at pH 7 and 5 ºC. H_2O/D_2O 90:10, 10 mM phosphate buffer, [oligonucleotide] =1 mM.

The H4 proton of tC^{o}2 shows NOE cross-peaks with its own H2'/H2" protons and H2'/H2" protons of previous cytosine in the sequence (C1) and exhibits some contacts with the 5'-5' intercalated cytosine within the C:C$^+$ stack. These last contacts correspond to the cytosine residue that forms the C:C$^+$ base pair with C7 that, according to the proposed topology, would be C15. Therefore, tC^{o}2 residue allowed to unequivocally assign all the cytosine residues and, consequently all the other residues in the sequence except for some of the thymine residues of the central poly-T linker.

Sequential H6/H8-H1'/H2'/H2" NOESY cross-peaks between non-exchangeable protons are observed for the segments: C1-tC^{o}2-G3-T4, C6-C7-G8-T9, C15-G16-T17 and C19-C20-G21-T22. Some characteristic H1'-H1' cross-peaks across the minor groove between intercalated residues are also observed confirming the following stacking order: tC^{o}2-C7 and C19-C15-C20. Other additional cross-

peaks are also found in this region: C7H5-C6H1', C14H5-T13H1' as well as those corresponding to the tC^{o}2 aromatic system (see *Figure* 6-15).

Regarding the exchangeable proton region of the spectra, the above mentioned protonated imino signals correspond to tC^{o}2:C20^{+} (15.00 ppm) and C7:C15^{+} (15.57 ppm) base pairs. In addition to expected imino-amino signals for these residues, an imino-imino contact is also observed between the hemiprotonated base pairs. More upfield, three guanine imino signals are observed in the region of WC-bonded imino protons (13.94,13.32 and 12.84 ppm). The signal at 13.94 ppm is the most intense and shows a small cross-peak with tC^{o}2H10, so it was assigned to G3. One of the two other imino signals was assigned to G8 (G8:C19 WC base pair). The other signal could not be unequivocally assigned. However, a G3H1-G21H1' contact is observed, indicating that, although formation of C6:G21 base cannot be asses, G3 and G21 face each other through the minor groove. Some contacts between thymine residues of the central loop, allowed to make a tentative assignment of these residues that are labelled in italics (see *Figure* 6-15). Chemical shifts are given *Table App* 14.

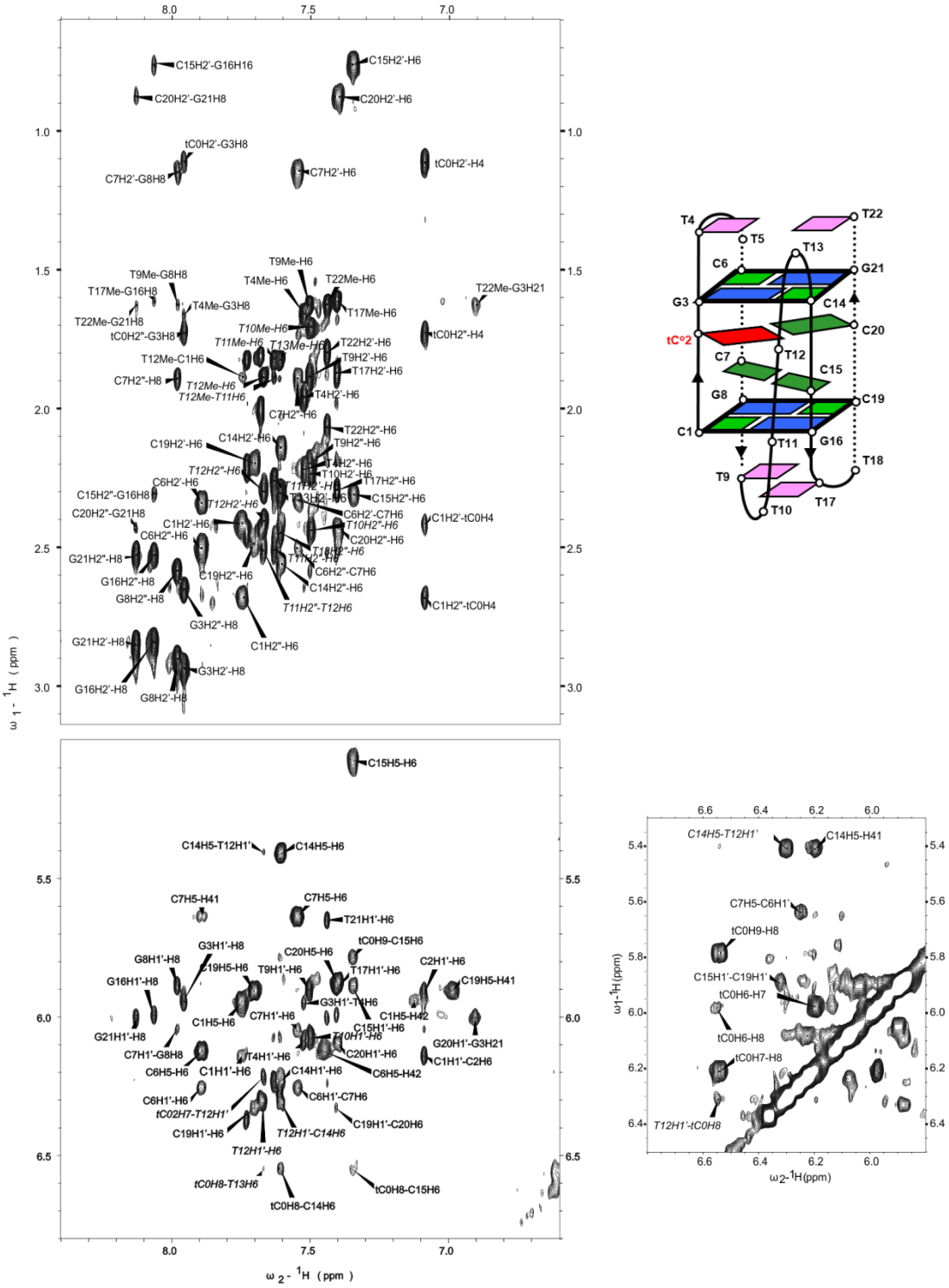


Figure 6-15. Non-exchangeable protons regions of NOESY spectrum (150 ms) of **NN4_tC°(2)** at pH 7 and 5 ºC. 10 mM phosphate buffer, [oligonucleotide] =1 mM. Schematic representation of the structure of **NN4_tC°(2)** at pH 7.

6.3.1.2. SOLUTION STRUCTURE OF **NN4_tC^o(2)**

The previously discussed assignment for the sequence **NN4_tC^o(2)** at pH 7 is consistent with the formation of a head-to-tail mini i-motif structure with one C:C$^+$ and one tC^o:C$^+$ base pair capped by not completely symmetric G:C:G:C minor groove tetrads (see *Figure* 6-15).

The complete assignment of the NOESY spectra of **NN4_tC^o(2)** at pH 7 was used for the calculation of its solution structure under these experimental conditions. A graphical view of the NOE contacts observed is shown in *Figure* 6-16.

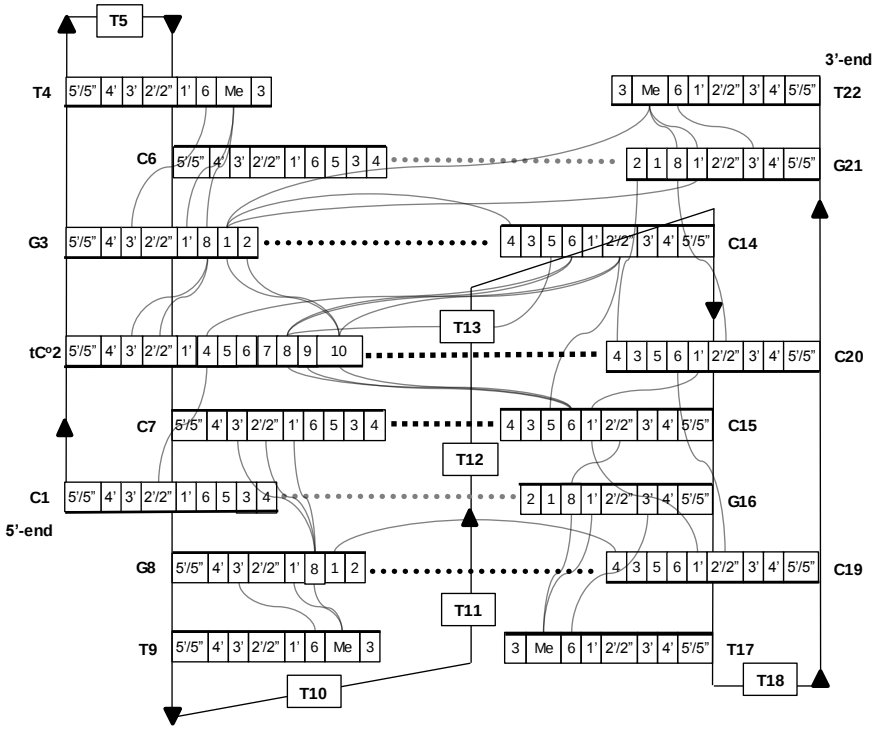


Figure 6-16. Scheme of the most relevant NOE contacts. Round dotted lines indicate G:C base pairs and square dotted lines stand for C:C$^+$ base pairs. C6:G21 and C1:G16 base pairs can be potentially formed but as characteristics cross-peaks have not been observed they are indicated in grey.

The ensemble of ten solution structures showed in *Figure* 6-17was successfully obtained with a RMSD value of 0.8 Å from approximately 115 experimental constraints (see *Table* 6-3 for further details on the calculation).

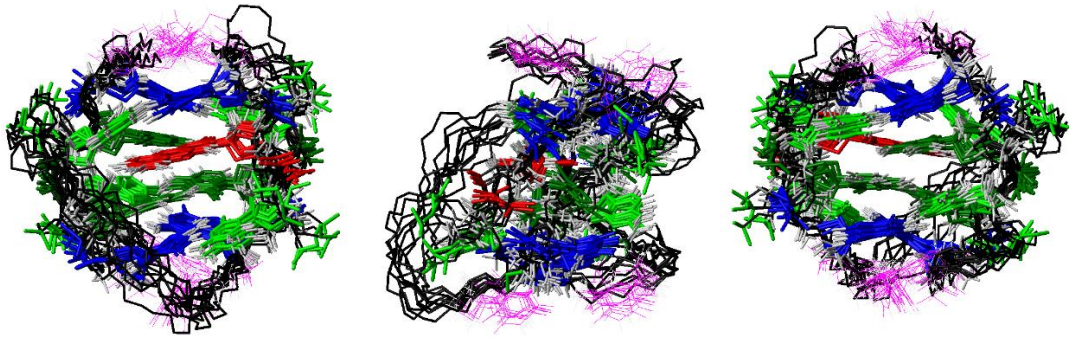


Figure 6-17. Different views of the ensemble of ten structures of **NN4_tC^o(2)**.

EXPERIMENTAL DISTANCE CONSTRAINTS		RMSD (Å)		AVERAGE RESIDUAL VIOLATIONS (AND RANGE)	
Total	115	Well-defined bases	0.5 ± 0.2	Sum of violation (Å)	6.35 (5.1-7.6)
Intra-residue	42	Well-defined heavy atoms	0.8 ± 0.3	Max. violation (Å)	0.68 (0.28-0.98)
Sequential	38	Backbone	1.8 ± 0.5	NOE energy (kcal/mol)	28 (22-34)
Range > 1	35	All heavy atoms	2.8 ± 0.7	Total energy (kcal/mol)	-1608 (-1478-1738)

Table 6-3. Experimental constraints and calculation statistics of **NN4_tC^o(2)**.

Calculated solution structure of **NN4_tC^o(2)** at neutral pH reveals the formation of a very similar construction to that obtained for **LL3**. The overall structure can be described as an i-motif structure with head-to-tail strand folding held together by the hydrogen bonds formed between one C:C$^+$, one tC^o:C$^+$ (see *Figure* 6-5, center) and two WC G:C base pairs. The formation of extra base pairs between the guanine residues, affords the assembling of two slipped G:C:G:C minor groove tetrads, which based on the experimental data seem to be not completely symmetric but are perfectly stacked on the hemiprotonated base pairs and thus, significantly contributes to the stabilization of the structure. However, for the calculation of the structure, the formation of symmetric G:C:G:C minor groove tetrads was forced, Very interestingly, the tC^o aromatic system is found perfectly fitted within the core of the structure, stacked under a WC G:C base pair (see *Figure* 6-18). The enhanced stacking interaction provided by bigger size of tC^o is most likely responsible for the remarkably enhanced stabilization exhibited by **NN4_tC^o(2)**. *Figure* 6-18 also reveals a general globular and compact shape of the structure, very similar to that observed for **LL3** and **NN4**.

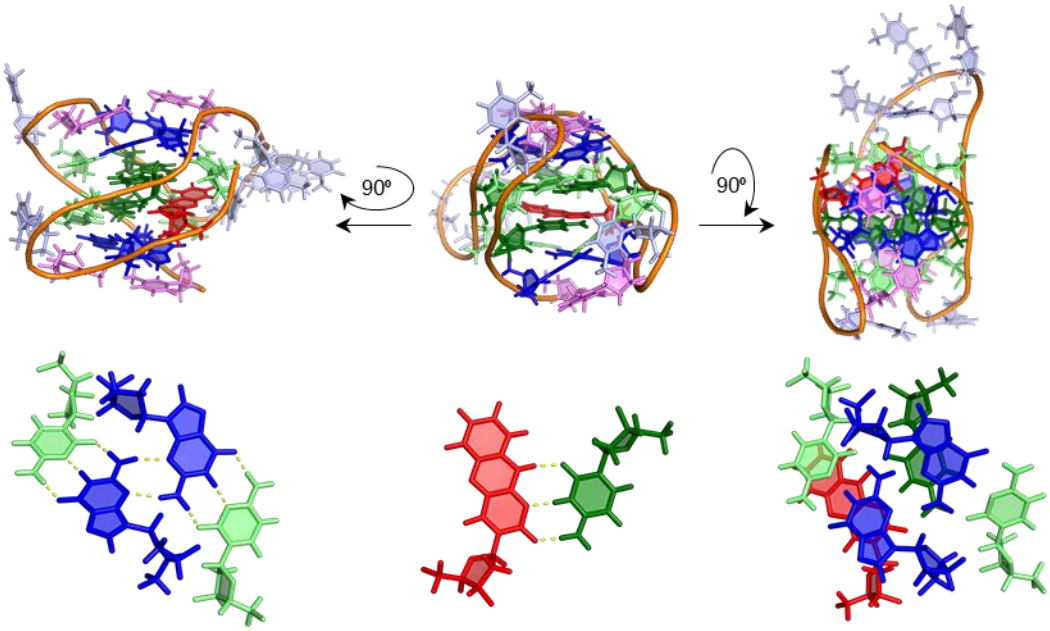


Figure 6-18. Top: Different stereoviews of the calculated structure of **NN4_tC^o(2)**. Bottom: Structural details of the calculated structure of **NN4_tC^o(2)**.

According to the dihedral angle values shown in *Table* 6-4, the backbone of the structure is very well-defined. Sugar conformations are shown in *Table* 6-5. Given the great similitude of this structure with that of **LL3**, a comparison between their structural parameters is more interesting than its individualized discussion. Therefore, a general analysis on this data points that, the described tendencies between residues observed for the structure of **LL3**, are loosely found in the dihedral angles values of the structure of **NN4_tC^o(2)**. This fact can be explained by the incorporation of tC^o within the structure, that might influence the configuration of the backbone due to its tricyclic nature and the loss of symmetry within the structure. Glycosidic dihedral angle values are, analogously to **LL3**, majorly in the anti range. Also observed for **LL3**, guanine residues exhibit chorally lower values.

Focusing now on the puckering of the sugars shown in *Table* 6-5, most of the residues, analogously to **LL3**, exhibit conformations in the C2'-*endo* region.

Nt	α		β		γ		δ		ε		ζ		χ	
	Avg.	OP	Avg.	OP	Avg.	OP	Avg.	OP	Avg.	OP	Avg.	OP	Avg.	OP
C1	-	-	169	0.7	107	0.2	123	1.0	-141	0.9	-	-	-144	1.0
tC°2	-19	0.4	-163	0.2	-93	0.3	120	1.0	-143	0.9	-97	0.8	-124	1.0
G3	-30	0.6	151	0.6	68	0.6	128	1.0	-130	0.9	-118	0.9	-104	0.9
T4	-136	0.5	145	0.6	144	0.6	150	1.0	-134	0.9	173	0.6	-172	1.0
T5	147	0.4	142	0.6	-167	0.2	128	1.0	-130	0.7	87	0.6	-167	1.0
C6	-146	0.6	180	0.7	100	0.7	124	1.0	-142	0.9	178	0.4	-114	1.0
C7	-23	0.6	-164	0.7	-59	0.4	103	1.0	-147	0.5	-103	0.9	-129	1.0
G8	22	0.4	-150	0.5	180	0.4	140	0.9	-132	0.9	-115	0.6	-109	1.0
T9	-176	0.9	-151	0.8	154	1.0	134	1.0	-139	0.9	180	0.9	-173	1.0
T10	21	0.0	166	0.7	180	0.4	136	0.9	-111	0.8	177	0.7	-158	0.6
T11	137	0.7	143	0.9	178	0.7	135	1.0	-130	0.8	-135	0.9	-168	0.6
T12	-160	0.4	-179	0.7	-176	0.7	136	1.0	-130	0.8	-126	0.9	-147	0.3
T13	-124	0.4	147	0.7	-180	0.4	128	1.0	-136	0.9	-169	0.7	-146	0.5
C14	156	0.4	-166	0.9	27	0.6	122	1.0	-150	1.0	127	0.7	-83	0.9
C15	83	0.5	-162	0.7	-135	0.8	74	0.8	-135	0.7	-103	0.8	-105	1.0
G16	-7	0.4	122	0.3	142	0.6	129	0.9	-111	0.9	-132	0.7	-129	0.9
T17	-110	0.6	-171	0.9	101	0.6	128	1.0	-103	0.9	173	0.8	-175	1.0
T18	-162	0.4	-164	0.8	172	0.3	145	1.0	-137	0.9	54	0.8	-162	1.0
C19	-165	0.6	163	0.5	31	0.8	120	1.0	-168	1.0	123	0.6	-40	1.0
C20	66	0.5	179	0.3	-176	0.5	53	0.7	-118	0.7	-111	0.9	-111	1.0
G21	-53	0.6	105	0.5	96	0.6	134	0.9	-107	0.9	-126	0.6	-104	1.0
T22	-180	0.6	-174	0.5	131	0.9	139	1.0	-	-	145	0.7	-175	1.0

Table 6-4. Average dihedral angles and order parameters of the structure of **NN4_tC°(2)** Average values correspond to the geometrical mean value and order parameters are use as indicators of angle definition: S=1, perfectly defined and S=0, random distribution.

Nt	Pseudorot.		Sugar Conformation
	Phase	Ampl.	
C1	157	32	C2'-*endo*
tC°2	146	48	C1'-*exo*, C2'-*endo*
G3	147	50	C1'-*exo*, C2'-*endo*
T4	206	44	C3'-*exo*
T5	166	45	C2'-*endo*
C6	179	39	C2'-*endo*, C3'-*exo*
C7	122	49	C1'-*exo*
G8	187	53	C2'-*endo*, C3'-*exo*
T9	227	48	C3'-*exo*, C4'-*endo*
T10	179	42	C2'-*endo*, C3'-*exo*
T11	164	44	C2'-*endo*
T12	171	47	C2'-*endo*
T13	172	41	C2'-*endo*
C14	110	29	O4'-*endo*, C1'-*exo*
C15	114	59	O4'-*endo*, C1'-*exo*
G16	163	49	C2'-*endo*
T17	187	27	C2'-*endo*, C3'-*exo*
T18	167	46	C2'-*endo*
C19	119	44	C1'-*exo*
C20	123	60	C1'-*exo*
G21	165	53	C2'-*endo*
T22	202	46	C3'-*exo*, C4'-*endo*

Table 6-5. Pseudorotation angle and amplitude values that allow the description of the puckering of the ribose ring in the calculated structure of **NN4_tC°(2)**.

6.3.1.3. THE ACIDIC SPECIES OF **NN4_tC^o(2)** IS ANALOGOUS TO THAT OF **NN4**

2D NMR spectra of **NN4_tC^o(2)** were acquired at several acidic conditions. At pH 6, the non-exchangeable protons region perfectly overlaps with signals at pH 7 and only small chemical shift changes are observed for some exchangeable protons, indicating that folding pattern and overall structure at pH 6 is the same as that at pH 7. As a general appreciation, guanine imino signals are better observed under these experimental conditions

At pH 5 new signals appear: up to 13 cytosine residues in the TOCSY spectrum and new amino-imino cross-peaks in the protonated imino protons region at 15 ppm in the NOESY spectrum. Two sets of tC^{o}2 signals are also observed, pointing to a slow equilibrium between two structured forms. One of the species is the one found at pH 7 and most of the cross-peaks found for this species are still observed. However, no signals corresponding to WC G:C base pairs formation are observed. At 5 ºC, both species are similarly populated. In order to fully characterize this acidic species, experimental pH was lowered to 4 and new NOESY spectra were recorded.

At pH 4, although some signals corresponding to the neutral species are still observed, the major species is the acidic one. In the imino protonated region, four signals are observed at 15.65, 15.24, 14.95 and 14.79 ppm. Three of the signals were identified to belong to the acidic form. The signal at 15.65 ppm exhibits cross-peaks with four couples of amino protons that correspond to C6:C14$^+$ and C1:C19$^+$ base pairs. The signal at 14.95 ppm corresponds to the tC^{o}2:C20+ base pair, as it shows cross-peaks with a couple of amino protons and the H10 proton of tC^{o}2. Finally, the signal at 15.24 ppm corresponds to a C7:C15$^+$ base pair (see *Figure* 6-19 left). The assignment of cytosine residues could be accomplished by the contacts of C20 and C14 residues with tC^{o}2 (C20H3$^+$-tC^{o}2H10, tC^{o}2H8-C14H6). The C1 residue could be assigned on the basis of characteristic C14H42-C1H2'/H2" protons between 3'-3' intercalated bases. C7H1'-C1H1' NOE contacts allowed distinguishing between C7 and C15. Two G:T imino-imino cross-peaks were observed. According to the observed contacts with guanine imino protons and amino protons of cytosine residues, these two pairs could be assigned to G8:T18 and G21:T5.

Therefore, the proposed structure of **NN4_tC^o(2)** at acidic pH is consistent with the formation of a head-to-tail i-motif structure with three C:C$^+$ and one tC^o:C$^+$ base pairs capped by G:T:G:T minor groove tetrads, quite similar to that observed for NN4 (*Figure* 6-19, right).

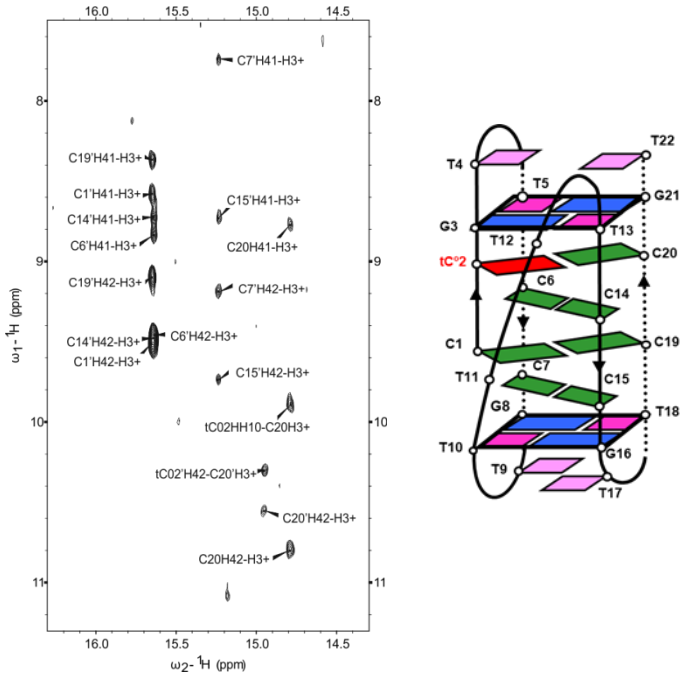


Figure 6-19. Hemiprotonated protons regions of NOESY spectrum (150 ms) of **NN4_tC°(2)** at pH 4 and 5 ºC. H_2O/D_2O 90:10, 10 mM phosphate buffer, [oligonucleotide] =1 mM (signals corresponding to acidic form are labelled with primes) (left). Schematic representation of the structure of **NN4_tC°(2)** at pH 4 (right).

6.3.2. NN4_tC°(6)

6.3.2.1. **NN4_tC°(6)** ALSO FOLDS INTO A MINI i-MOTIF STRUCTURE AT NEUTRAL pH

At neutral pH, seven H5-H6 cross-peaks and ten Me-H6 are observed in the TOCSY spectrum, indicating the formation of a unique species. Moreover, only one aromatic spin system corresponding to one tC° residue is observed.

Four cytosine residues are involved in hemiprotonated C:C$^+$ base pairs. Two nearly overlapped protonated imino signals are observed at 15.40 and 15.38 ppm. Each of these signals exhibits cross-peaks with two couples of amino protons. These four cytosines could be easily assigned to C2, C7, C15 and C20 on the basis of sequential GH8-CH2'/H2" cross-peaks with guanine residues. Therefore, as expected, tC° residue is not involved in a hemiprotonated base pair (see *Figure* 6-20).

tC°6 residue exhibits sequential C7H6-tC°6H2'/H2" and tC°6H4-T5H1'/H2'/H2" cross-peaks and, interestingly, some cross-peaks with a cytosine residue not involved in hemiprotonated base pairs. Stacking connections H6/H8-H2'/H2" could be stablished for four 5'-CCGT-3' tracts (*Figure* 6-21). Although not as clear as for **NN4_tC°(2)**, the presence of some cross-peaks between thymine

connecting loop residues and different cytosine and thymine residues, pointed to a head-to-tail like folding, with the thymine-loop located near the major groove.

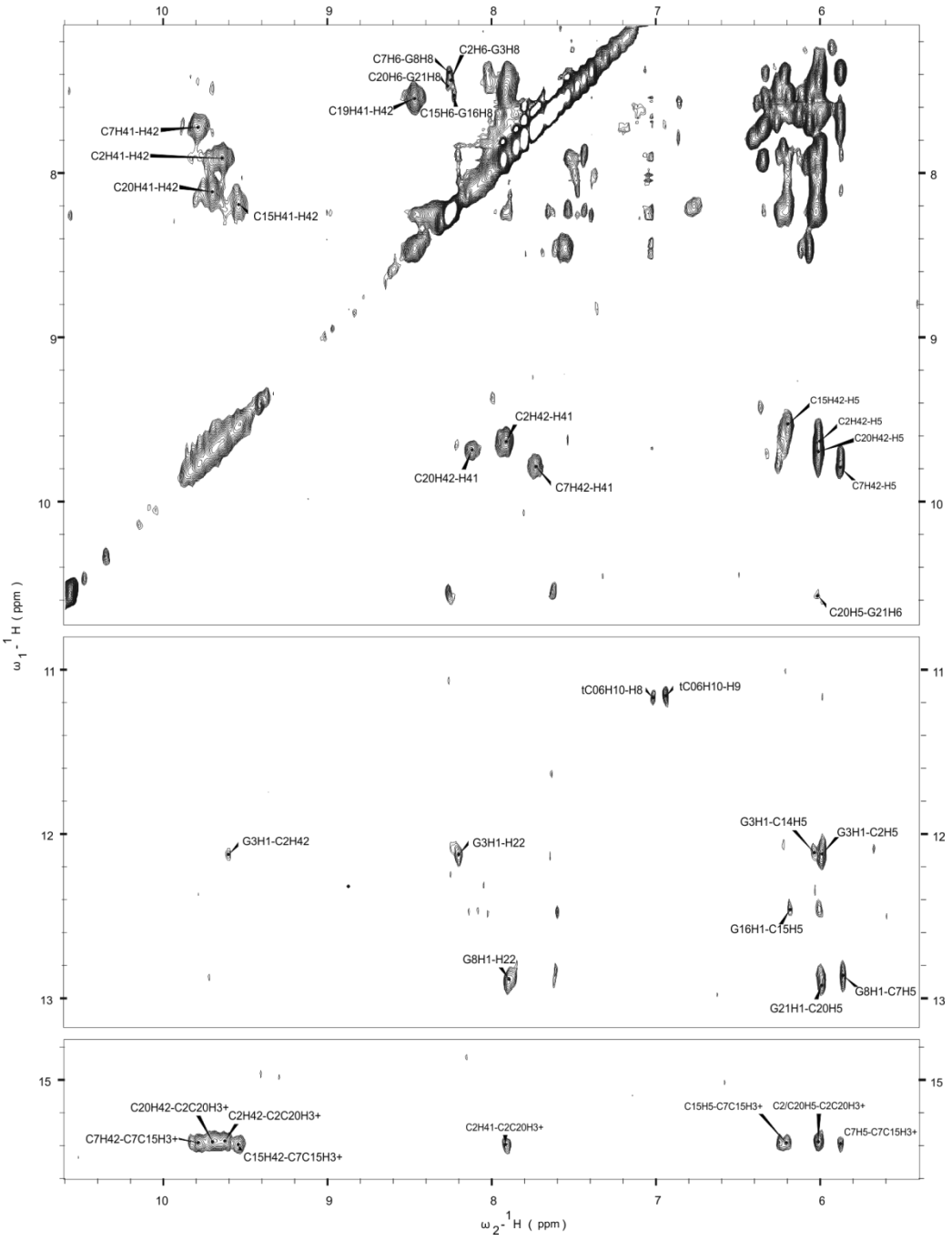


Figure 6-20. Exchangeable protons regions of NOESY spectrum (150 ms) of **NN4_tC°(6)** at pH 7 and 5 ºC. H_2O/D_2O 90:10, 10 mM phosphate buffer, [oligonucleotide] =1 mM.

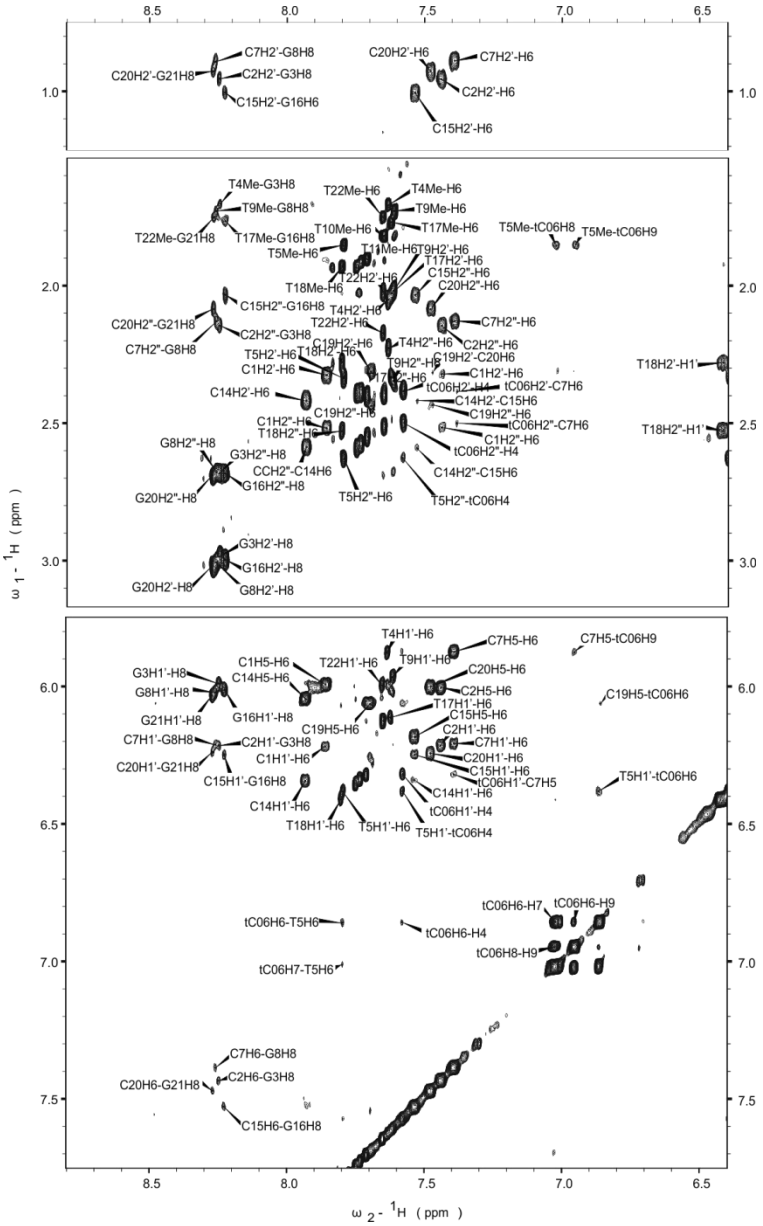


Figure 6-21. Non-exchangeable protons regions of NOESY spectrum (150 ms) of **NN4_tC°(6)** at pH 7 and 5 ºC. H_2O/D_2O 90:10, 10 mM phosphate buffer, [oligonucleotide] =1 mM.

Thus, assuming a head-to-tail folding and considering the observed stacking connections of the fragment T5-tC°6-C7-G8-T9, the assignment of the rest of the residues could be carried out. Considering $C7{:}C15^+$ and $C2{:}C20^+$ hemiprotonated base pairs formation, the following connections were established: C7-C15, C14-C15-G16-T17. At this point the cytosine residue showing cross-peaks with

tC^{o}6 and not involved in the formation of hemiprotonated base pairs could be identified as C19. From C19, another 5'-CCGT-3' fragment could be assigned: C19-C20-G21-T22. Finally, the remaining fragment C1-C2-G3-T4, was assigned based on the formation of C2:C20$^+$ base pair.

A complete assignment was achieved, although thymine residues of the central connecting loop were tentatively assigned. Chemical shifts are given in *Table App* 15.

Characteristic G:C Watson-Crick cross-peaks exhibit very low intensity (in fact, only observed at one side of the diagonal). G3:C14, G8:C19 and G16:C1 could be assigned based on stacking cross-peaks of guanine residues with C2, C7 and C15, respectively (see *Figure* 6-20). Although no cross-peak was observed between any guanine imino proton and the H10 proton of tC^{o}6, the chemical shift of H10 (11.17 ppm) indicates its implication in hydrogen bond formation.

Altogether, at neutral pH the NMR data is consistent with the formation of a unique species represented in *Figure* 6-22. Such structure consists of a head-to-tail mini i-motif structure capped by G:C WC base pairs most likely assembled forming a slipped tetrad. However, the poor quality of the cross-peaks involving exchangeable guanine imino protons prevented its full characterization. The fluorescent tC^o unit would be located in one of these tetrads.

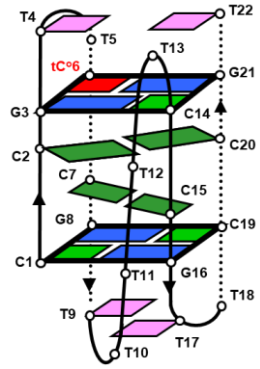


Figure 6-22. Schematic representation of the structure of **NN4_tC^o(6)** at pH 7.

6.3.2.2. tC^o IS DRIVEN TO OCCUPY CENTRAL POSITIONS IN THE ACIDIC SPECIES OF **NN4_tC^o(6)**

Under acidic experimental conditions, 2D NMR experiments reveal the presence of an alternative species and two sets of aromatic signals corresponding to tC^o residues are observed. According to non-exchangeable protons, both species seem to qualitatively just as populated. However, most of the exchangeable proton signals correspond to the acidic species. It can be clearly observed that for this acidic species, tC^{o}6 is forming a hemiprotonated base pair and located in the center of the stack (H10 proton of tC^o exhibit characteristic amino-H2'/H2" cross-peaks between 3'-3' intercalated base pairs alongside the major groove). Starting from tC^{o}6 signals, C14, C1 and C19 could be specifically assigned.

Imino-imino cross-peaks between intercalated pairs (see *Figure* 6-23, left) allowed to identify the C7:C15$^+$ hemirprotonated pair, although due to the overlapping exhibit by their NOE signals, C7 and C15 could not be specifically assigned. C20 residue was tentatively assigned based on an unusual H5 chemical shift (5.34 ppm): As the similar effect can be observed for C19 (5.57 ppm), we propose that this upshifting is due to the shielding provoked by the interaction across the major groove with the aromatic system of tC^{o}6, located in the middle of the two base pairs in which C20 and C19 are involved.

Four TH3-GH1 cross-peaks are also observable, supporting to the formation of G:T:G:T tetrads. G3 and G21 residues were specifically assigned. Extra experiments in order to achieve a better signal dispersion in the aromatic region are required to specifically label C7 and C15 and complete the sequential assignment. However, the described experimental data is consistent with the formation of a similar i-motif structure found for **NN4** at acidic pH (*Figure* 6-23, right).

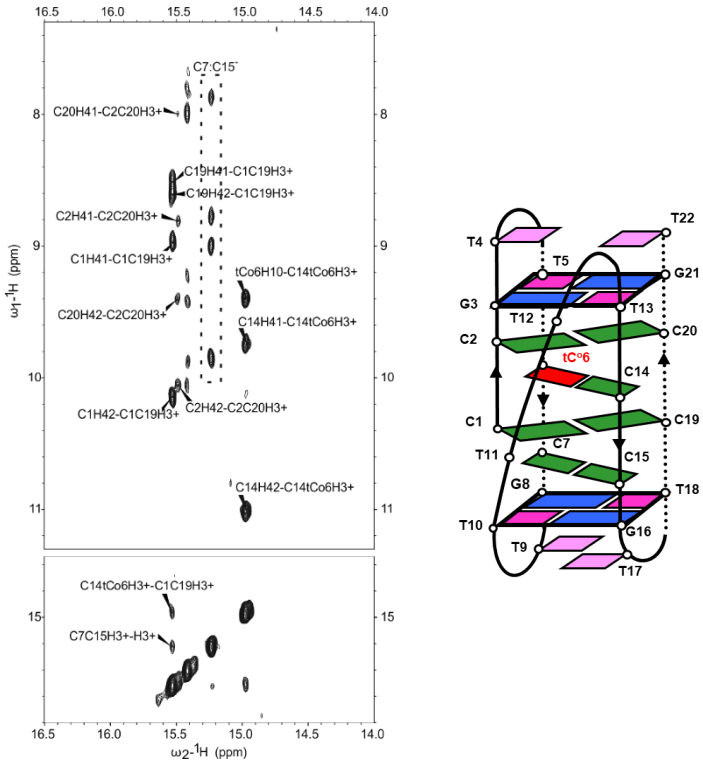


Figure 6-23. Hemiprotonated protons regions of NOESY spectrum (150 ms) of **NN4_tC°(6)** at pH 5 and 5 ºC. H_2O/D_2O 90:10, 10 mM phosphate buffer, [oligonucleotide] =1 mM (left). Schematic representation of the structure of **NN4_tC°(6)** at pH 5 (right).

6.4. FLUORESCENCE AND CD-MONITORED pH-DRIVEN TRANSITIONS

Once the structures of the i-motif structures involved in the pH-dependent structural equilibrium are well-established, a more detailed study based on CD and fluorescence spectroscopy was performed. We also aimed to test the potential of tC^o as a local fluorescent probe. We have focused on the estimation of the pH_T values for each transition calculated from both, CD and fluorometry experiments. The obtained results are summarized in *Table* 6.

NAME	CD 265 nm		CD 295 nm		FLUOR.
	pH_{T1}	pH_{T2}	pH_{T1}	pH_{T2}	pH_T
NN4	6.1 ± 0.1	8.0 ± 0.1	6.1 ± 0.1	7.9 ± 0.1	-
NN4_tC^o(2)	5.4 ± 0.1	8.3 ± 0.1	-	8.3 ± 0.1	8.0 ± 0.1
NN4_tC^o(6)	-	7.9 ± 0.1	6.0 ± 0.1	7.8 ± 0.1	6.3 ± 0.1

Table 6-6. Summary of the pH_T values calculated from CD and fluorescence experiments for **NN4**, **NN4_tC^o(2)** and **NN4_tC^o(6)**. 25 mM phosphate buffer, [oligonucleotide] = 2 μM (CD) 0.2 μM (Fluorescence).

6.4.1. CD

Series of CD spectra were recorded over a wide range of pH for both tC^o-containing sequences (see *Figure* 6-24). Spectral features were already described in Section 2, highlighting the band observed at 300 nm at acidic pH values. From the superposition of these spectra for each sequence, two isosbestic points (250 and 285 nm) indicate two structural transitions involving three well-defined species. According to the previously presented NMR results, these species must be the acidic form, the neutral form and the unfolded structure. Each species exhibits a different wavelength for its ellipticity maximum (295, 265 and 275-280 nm, respectively) that is observable in the range of pH in which these species become predominant. Remarkably, the set of spectra recorded for **NN4_tC^o(2)**, reveal a less defined band at 295 nm in comparison to the other **N**-derived sequences. This information is in tight agreement with the NMR results, that confirm that the formation of the acidic species for NN4_tCo2 is not so favorable, and that even at low pH both neutral and acidic forms are observed.

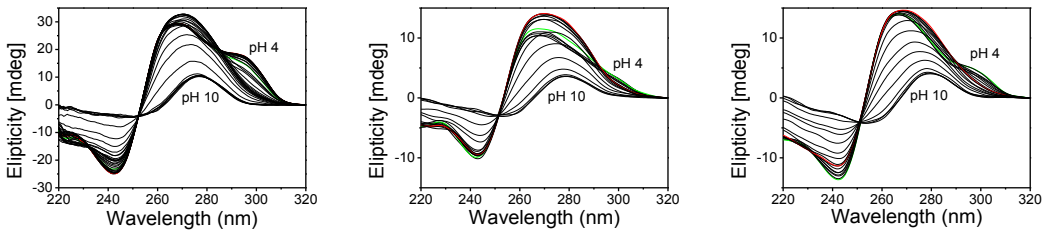


Figure 6-24. CD spectra superposition of **NN4** (left), **NN4_tC°(2)** (center) and **NN4_tC°(6)** (right) at 5 ºC and different pH values. 25 mM phosphate buffer, [oligonucleotide] = 2.0 µM. Spectra at neutral pH are shown in red and at pH 5 in green.

pH-titration curves were obtained by plotting maximum ellipticity values at the two different wavelengths (265 and 295 nm) attributable to the acidic and neutral species as a function of pH (see *Figure* 6-25 and *Figure* 6-26). First, it is worth to mention that the relative intensity of the ellipticity signal does not reveal any substantial information since it depends very specifically on the structure under study. Thus, analogous structures with different sequences (tC° in different positions, in this case) might exhibit differences in the ellipticity intensity. However, what really makes these experiments so useful and interesting is to compare the pH values at which do the structural transitions occur. At 265 nm, a first transition (represented in green) occurs between pH 5-6 from lower to higher ellipticity values, that might correspond to the interconversion of the two i-motif structures. The second transition occurs at around pH 8 and may correspond to the denaturalization of the structures (represented in red). The variation of ellipticity notable decrease indicating the loss of structuration. Interestingly, from a qualitative point of view, it can be observed that the neutral species of **NN4_tC°(2)** remains as the prevalent species for a wider pH range than for **NN4** and **NN4_tC°(6)**.

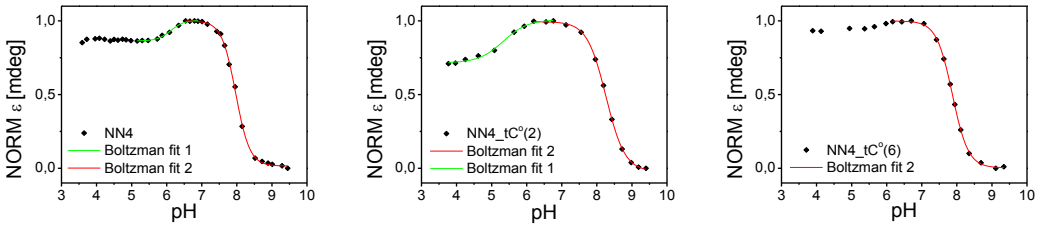


Figure 6-25. CD-monitored pH titration curves at 265 nm of **NN4** (left), **NN4_tC°(2)** (center) and **NN4_tC°(6)** (right) at 5 ºC. 25 mM phosphate buffer, [oligonucleotide] = 2.0 µM. Boltzmann fit represented in red for the denaturalization of the structures (pH_{T2}) and in green for the equilibrium between species (pH_{T1}).

Remarkably, at 295 nm one can follow the same transitions described by the plots at 265 nm, plus a third transition at around pH 4 (*Figure* 6-26, close to the pK_a value of cytosine) that might correspond to the denaturation of the acidic species due to the complete protonation of cytosine residues.

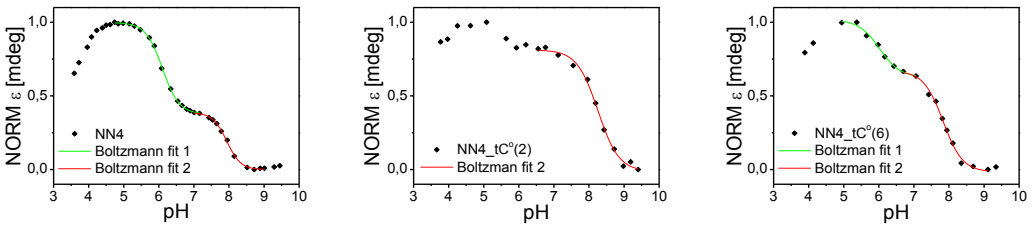


Figure 6-26. CD-monitored pH titration curves at 295 nm of **NN4** (left), **NN4_tC°(2)** (center) and **NN4_tC°(6)** (right) at 5 ºC. 25 mM phosphate buffer, [oligonucleotide] = 2.0 μM. Boltzmann fit represented in red for the denaturalization of the structures (pH_{T2}) and in green for the equilibrium between species (pH_{T1}).

From the representations shown in *Figure* 6-26 and 27, quantitative pH_T could be calculated applying a Boltzmann sigmoidal fit. *Table* 6-6 summarizes these results, being pH_{T1} values the ones calculated for the transition between the acidic and the neutral i-motif species and pH_{T2} values corresponding to the total denaturation of the sequences. Due to the low difference in ellipticity and the fewer points available for the fitting, pH_{T1} at 295 nm for **NN4_tC°(2)** and pH_{T1} at 265 nm for **NN4_tC°(6)** could not be calculated. Notably, the results obtained at different wavelengths perfectly match.

Very interestingly, the pH_{T1} value obtained for **NN4_tC°(2)** (5.4) is significantly lower than that obtained for **NN4** (6.1). This quantitative information reaffirms the prevalence of the neutral species of **NN4_tC°(2)**, that can be explained based on the very favorable stacking interactions displayed by the tC° residue in its structure that confer significantly enhanced thermal stability to the motif. This stability enhancement is translated to a transition from the acidic to the neutral form at lower pH values. **NN4_tC°(6)** displays values pH_{T1} values practically equal to those calculated for **NN4**.

Focusing now on the transition from the structured neutral species to the completely unfolded oligonucleotide, a similar trend is observed confirming the previously postulated analysis. pH_{T2} values of **NN4** and **NN4_tC°(6)** are practically equal (8.0 and 7.9, respectively) and remarkably high for i-motif structures. The relative increase observed for **NN4_tC°(2)** (pH_{T2} = 8.3) is again related to the high stability of the structure provided by the tC° residue which, as a conclusion, allows the observation of the mini i-motif structure of **NN4_tC°(2)** at a wider range of pH. *Figure* 6-27 shows a schematic representation of the transitions observed for this sequence.

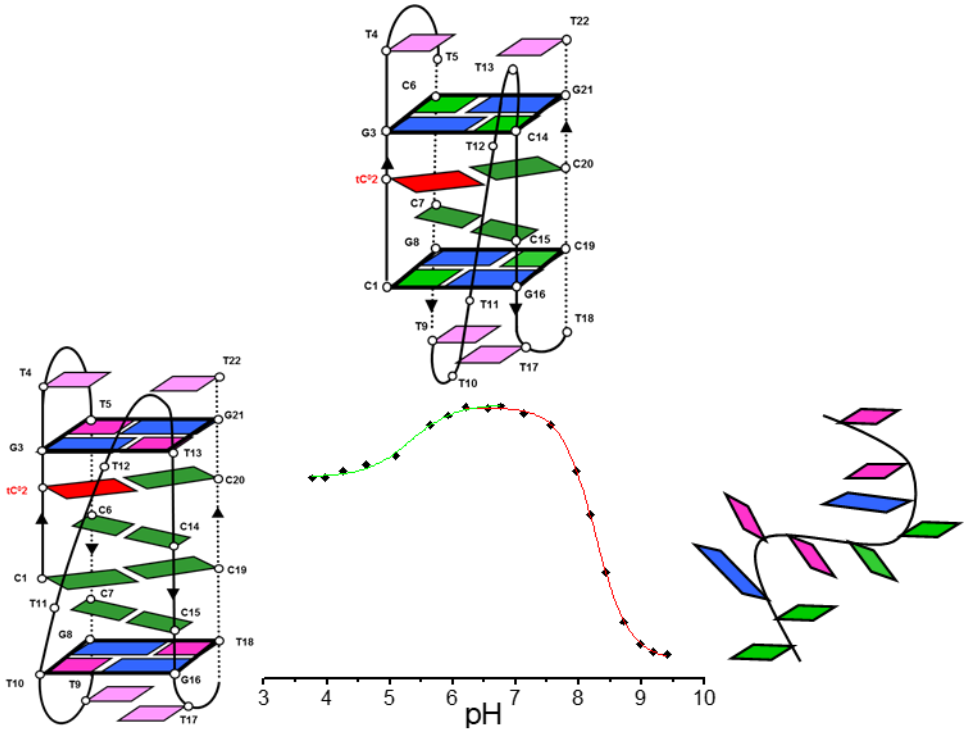


Figure 6-27. Schematic representation of the structural transition observed by CD spectroscopy for **NN4_tC°(2)**.

6.4.2. FLUORESCENCE

pH-titration experiments were analogously monitored by fluorometry. We expected that the environmental changes suffered by tC° moieties within the i-motif structures of **NN4_tC°(2)** and **NN4_tC°(6)** to be traceable by fluorescence spectroscopy.

In *Figure* 6-28, excitation spectra at a wide range of pH values are shown overlapped for the sequences **NN4_tC°(2)** and **NN4_tC°(6)**. The first remarkable information is that under acidic experimental conditions, both sequences exhibit a high quenching of the fluorescence signal. This is consistent to the NMR studies that reveal that, in both cases, at pH 4, tC° residue is predominantly involved in hemiprotonated base pairs. Upon basification, the quenching of the signal gradually diminishes indicating the disruption of the hemiprotonated base pairs and thus, the overall i-motif structure. However, recovery of the fluorescence signal occurs at very notably lower pH values for **NN4_tC°(6)** compared to **NN4_tC°(2)**. Interestingly, a unique transition is predominantly observed for both sequences.

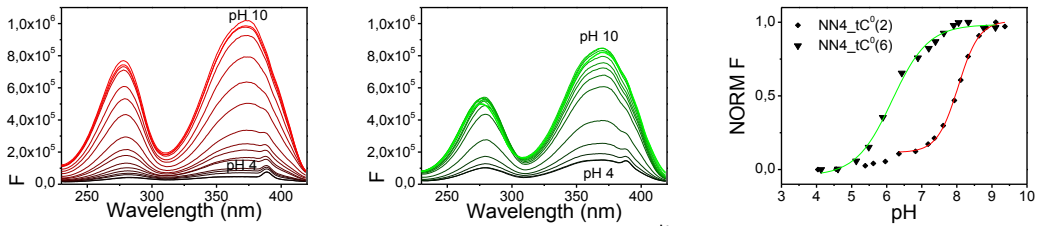


Figure 6-28. Fluorescence excitation spectra recorded at different pH values for **NN4_tC°(2)** (left) and **NN4_tC°(6)** (center). Fluorescence-monitored pH-titration curves for **NN4_tC°(2)** and **NN4_tC°(6)** (right). Boltzmann fit represented in green for the sequence **NN4_tC°(6)** and in red for the sequence for **NN4_tC°(2)**. 5ºC. 25 mM phosphate buffer, [oligonucleotide] = 0.2 µM.

Determination of pH_T values helped to analyze the structural transition being observed by fluorescence in each case. *Figure* 6-28*, right)* shows the experimental data at 370 nm plotted as a function of pH and the corresponding Boltzmann fits. For **NN4_tC°(2)**, in both i-motif structures tC°2 residue is involved in hemiprotonated base pairs and stacking interactions are also maintained. However, an important change in the chemical environment of the fluorophore occurs upon unfolding. The fluorescence unfolding profile and the high pH_T value obtained for **NN4_tC°(2)** (pH_T = 8.0, red), reveals that no significant changes in fluorescence occur for the interconversion between the two species and only the unfolding process is mainly observed by fluorescence. Small changes in fluorescence observed at low pH (see *Figure* 6-28*,* left*,* between pH 5 and 7) were not adjusted, but seem to coincide to a similar value than that obtained by CD (*Table* 6-6)

On the other hand, obtained pH_T value for **NN4_tC°(6)** is considerably lower (pH_T = 6.1, green). According to the proposed structures at neutral and acidic pH, transition between the two i-motifs implies a dramatic change in the chemical environment of tC°6 residue. At neutral pH, it is involved in a WC base pair and the fluorescence is expected to be no so different from that of the random form (see *Figure* 6-10). However, upon acidification, this WC base pair is disrupted and tC°6 forms an hemiprotonated base pair. The interconversion between the two i-motif structures is indeed sensitive to fluorescence. Thus, calculated pH_T, in this case corresponds to this transition rather than the unfolding process. In terms of quenching, the transition from forming a $tC^o:C^+$ base pair to a G:tC° base pair dramatically unquenches the fluorescence performance of tC°.

6.5. EXPLORING THE USE OF tC° FOR THE VISUALIZATION OF i-MOTIFS *IN VIVO*

The good results obtained for the structure of **NN4_tC°(2)** in terms of stability and fluorescence behavior, made us think of this sequence as a very strong candidate to test the visualization of the mini i-motif structure *in vivo*. Heartened by the bioinformatics analysis results presented in the previous chapter that pointed that mini i-motif-forming sequences are prevalent in the human genome, the next step in order

to prove the biological relevance of these peculiar structures is to challenge their formation in a cellular media.

6.5.1. tC^o SIGNAL IS NOT QUENCHED UPON DUPLEX FORMATION

Before moving to microscopy studies, a preliminary consideration had to be tested to validate the use of tC^o as a specific fluorescent probe for the formation of i-motif structures. As it is described in the literature, one of the substantial benefits exhibited by tC^o is that shows a notably low quenching of the emitted fluorescence signal when hybridized in a B-DNA duplex context.

In order to asses if the difference in quenching displayed by tC^o in an i-motif or a duplex structure is enough for the discrimination of the formation of both structures, fluorescence experiments were recorded for **NN4_tC^o(2)** in the presence of its complementary strand (d(ACGGAACGGAAAAACGGAACGG)) at different temperatures.

As it can be observed in *Figure* 6-29, in the presence of its complementary sequence, the resulting duplex exhibits a very low quenching of the fluorescence signal provided by tC^o moiety at any temperature. Interestingly, as temperature increases, the fluorescence signal becomes more quenched. On the other hand, as it was presented in *Section* 2, the quenching observed for **NN4_tC^o(2)** alone when forming a mini i-motif structure is almost total.

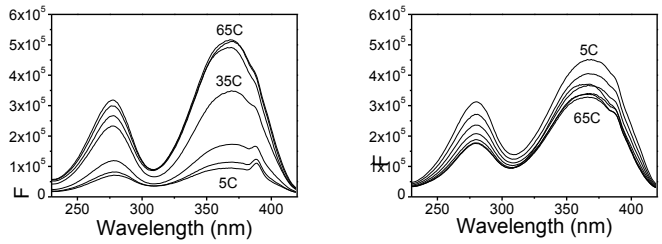


Figure 6-29. Fluorescence excitation spectra recorded at different temperature for **NN4_tC^o(2)** (left) and **NN4_tC^o(2)** in presence of its complementary sequence (right) at pH 7. 25 mM phosphate buffer, [oligonucleotide] = 0.2 μM.

6.5.2. THE FOLDING OF THE STRUCTURES CAN BE FOLLOWED BY LIVE-CELL MICROSCOPY

In order to asses if the formation of mini i-motif structures is feasible in the cell nucleus, HeLa cells (incubated at 37 ºC) were transfected with **NN4_tC^o(2)** in the presence of Lipofectamine 3000 ([oligonucleotide] = 500 nM). As a negative control, a shorter sequence containing tC^o that cannot adopt a folded i-motif structure under physiological conditions (d(CCGTTtC^{o}CGT)) was used. The fluorescent oligonucleotides were visualized by live-cell fluorescence microscopy 17 hours after transfection. These

experiments were performed in collaboration with M. Terrazas and A. Gandioso from Prof. M. Orozco's group in the IRB Barcelona facilities.

As evidenced in *Figure* 6-30, **NN4_tC^o(2)** entered the cells and spontaneously localized to the cell nucleus (see co-localization of cyan, corresponding to tC^o emission, and red dyes in cells with nuclei stained with DRAQ5 dye (far-red emission)). In the case of cells treated with the negative control, a strong cyan fluorescence intensity was observed, which was significantly higher than that observed in cells treated with **NN4_tC^o(2)** as shown in *Figure* 6-31. These results suggest that the negative control remains completely unfolded inside the cells, as expected. In contrast, **NN4_tC^o(2)** might be present in both folded and unfolded forms. This partial decrease in fluorescence in respect to the negative control, is in perfect agreement with the thermal denaturation studies, which confirmed a T_m value for **NN4_tC^o(2)** of 38.8 ºC, reaffirming that at 37 ºC, approximately 50% of the oligonucleotide is folded.

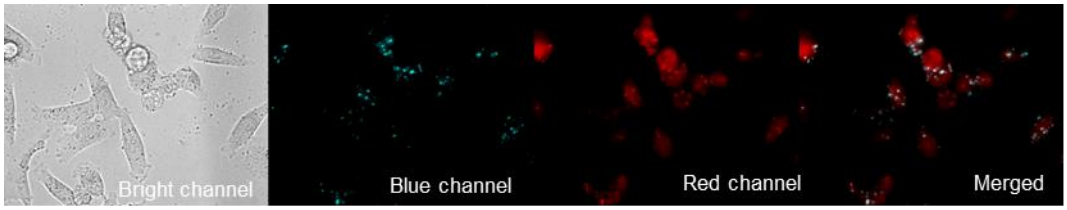


Figure 6-30. Sequence of images showing the co-localization of **NN4_tC^o(2)** in the nuclei of HeLa cells using DRAQ5 as dye under physiological conditions.

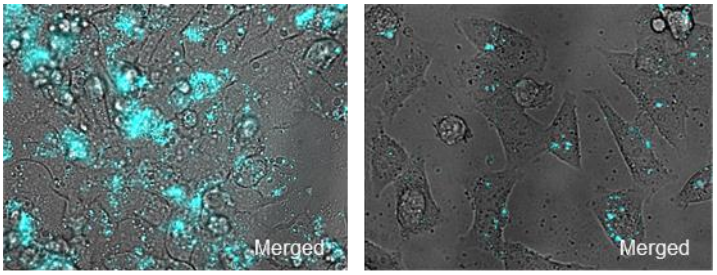


Figure 6-31. Confocal microscopy images of HeLa cells transfected with the negative control (left) and NN4_tC^o(2) (right) under physiological conditions.

Encouraged by these observations, we next proceeded to induce the complete folding of **NN4_tC^o(2)** by acidification of the cellular media. As observed in *Figure* 6-32, addition of increasing amounts of citric acid to cells transfected with **NN4_tC^o(2)** led to a gradual decrease in fluorescence intensity. A first addition of citric acid (final concentration 3 mM) already translates into a slight decrease in the fluorescence emission. Moreover, a second addition of citric acid (final concentration 17 mM), provokes the almost complete quenching of the fluorescence as a result of the folding of the motif.

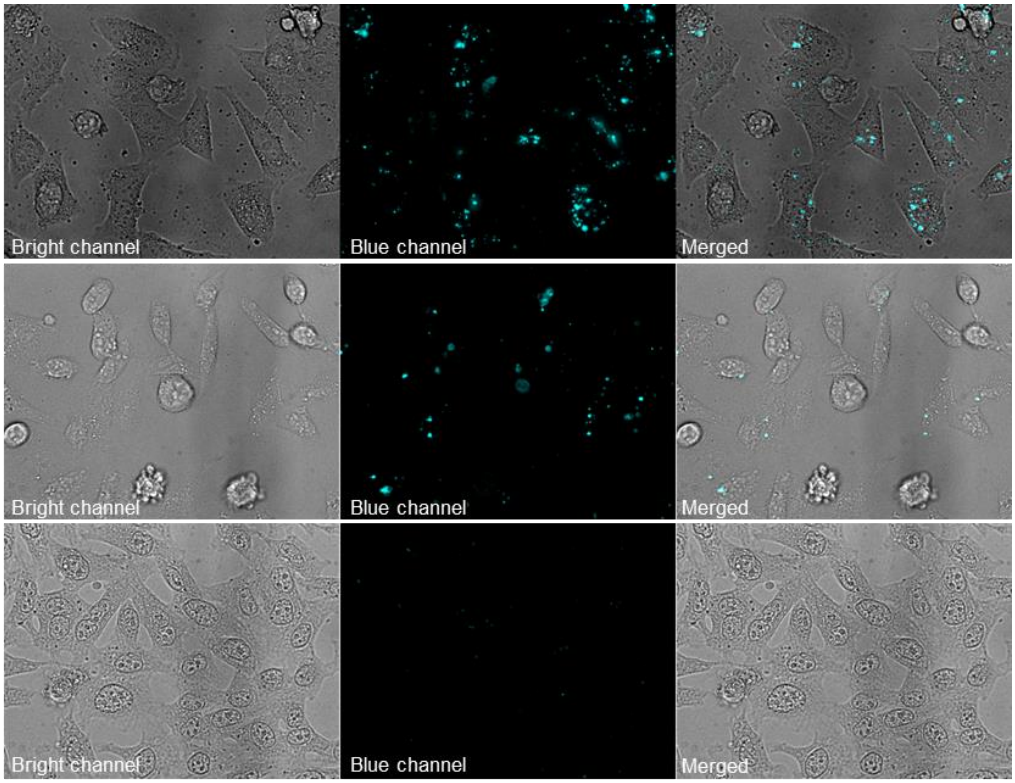


Figure 6-32. Sequence of images showing the quenching of the fluorescence signal of NN4_tC^o(2) upon acidification of transfected Hela cells. Top: physiological conditions. Center: 3 mM citric acid. Bottom: 17 mM citric acid.

Very remarkably, these results show that the formation of mini i-motif structures in the cell nucleus and the transition from unfolded to folded forms upon pH modulations, can be monitored by fluorescence microscopy. Furthermore, to our knowledge, this is the first report on the utilization of a fluorescent nucleobase for the visualization of i-motif structures *in vivo*. Moreover, tC^o or other strategically chosen fluorophores incorporated in mini i-motif structures might be able to build nanotechnological pH sensors in a range of pHs normally inaccessible to i-motif based sensors.

6.6. CONCLUSIONS

tC^o-containing sequences were successfully synthesized by standard SPOS methodology. Both, **NN4_tC^o(2)** and **NN4_tC^o(6)** unambiguously show stable i-motif formation at neutral pH as assessed by NMR and CD spectroscopy, which reveals that the incorporation of tC^o does not impede the formation of i-motif structures for the chosen sequences and maintain the great stability values observed for **NN4**.

The complete characterization of the structures was achieved based on NOESY experiments, concluding that both tC^{o}-containing sequences exhibit the same pH-dependent equilibrium between different i-motif species than that observed for **NN4**. Notably, the neutral form of **NN4_tC^{o}(2)** is significantly favored over the acidic species for this sequence.

The solution structure of **NN4_tC^{o}(2)** was well-defined on the basis of restrained molecular dynamics, showing very similar features to other mini i-motif structures. Most remarkably, the formation of tC^{o}:C^{+} base pairs accommodate perfectly within the structure, showing an enlarged stacking surface with G:C capping base pairs.

CD and fluorescence studies endorsed the use of tC^{o} as an advantageous fluorescent probe, that allowed the characterization of very precise structural transitions within the motifs. Very interestingly, tC^{o} fluorescence signal is completely quenched when the residue is hemiprotonated, while its fluorescence performance is slightly modified when hybridized in WC-like G:tC^{o} base pairs.

Finally, **NN4_tC^{o}(2)** was successfully introduced in HeLa cells. Confocal fluorescence microscopy experiments confirmed that the sequence is localized in the cell nucleus and that under physiological conditions, its fluorescent emission is quenched in comparison to a negative control. Furthermore, the folding equilibrium of the sequence can be monitored upon pH regulation of the cellular media. These novel results encourage the use of tC^{o} as a fluorescent nucleobase for the visualization of i-motif structures *in vivo*.

7. SYNTHESES

7.1. OLIGONUCLEOTIDE SYNTHESIS

The theoretical approach behind the synthesis of nucleic acids can be easily described as the formation of phosphodiester linkages between the 3'-OH group of one nucleoside with the 5'-OH group of another nucleoside. Regardless of its simplicity, the synthesis of nucleic acids is not so straightforwardly achieved since the selectivity of the reaction and the reactivity of the functional groups involved in the process must be taken under consideration.

As a general strategy, 3'-derivatized nucleosides are used together with different coupling agents to increase their reactivity. Historically, the first described methods for the synthesis of nucleic acids used 3'-phosphorilated nucleosides and coupling agents such as DCC or MSNT. However, the low reactivity of the phosphate group afforded low reaction yields and therefore the methodology was improved. The most important breakthrough in this field raised from the use of phosphoramidites as building blocks for the synthesis. The advantage behind these derivatized nucleosides is that they include a very reactive yet stable P(III) species that is easily activated by an acidic catalyst that, after reacting, can be straightforwardly oxidized to the natural P(V) form.[251,252] Further details that define the high efficiency of this methodology include the use of tetrazole derivatives as coupling agents, since their acidic nature is not strong enough to deprotect the 5'-OH group, the capping step before the oxidation that prevents the extension of side reactions and the straightforward cleavage and deprotection process in basic media. The orthogonal protecting approach for the nucleobases in respect to the 5'-OH group, uses standard protecting groups labile upon basic treatment: dA^{Bz}, dC^{Bz}, dG^{iBu}.

Another remarkable step forward for the synthesis of oligonucleotides was the automation of the method. The use of automatic synthesizers represents notable time and resources savings that translate into a huge increase in productivity. Generally, a trustable 98-99% yield is expected for each coupling step and the synthesizer allows checking if these quantitative yields are obtained by an indirect quantification of the DMT groups after the deprotection step. Automatic synthesizers take advantage on the use of solid supports and allow programming personalized synthetic cycles. The resins used as solid supports are typically functionalized with an inert spacer and a functional group with first nucleotide of the 3'-end of the chain attached and properly protected.

7.1.1. SYNTHESIS OF LINEAR OLIGONUCLEOTIDES

All the oligonucleotides synthesized in this thesis were prepared using the phosphoramidite strategy described above (phosphite triester method). Commercial 3'-(2-cyanoethyl)-*N*,*N*-diisopropyl-phosphoramidites of 5'-*O*-(4,4'-dimethoxytrityl)-deoxyribonucleosides were used for the syntheses, which were performed over controlled pore glass (CPG) beads functionalized with the 5'-*O*-DMT-nucleoside of the 3'-end of the oligonucleotide through a succinyl group. The general synthetic cycle is graphically represented in *Figure* 7-1 and key steps are listed and described hereafter:

- Detritylation: initial deprotection of the 5'-OH group of the 3'- end nucleotide by removal of the 4,4'-dimethoxytrityl (DMT) group upon acidic treatment with trichloroacetic acid (TCA).
- Coupling: addition of the properly protected phosphoramidite over the free 5'-OH group of the 3'-end nucleotide prior *in situ* activation by tetrazole or BTT.
- Capping: acetylation of the free 5'-OH groups that did not react in the previous step in order to prevent obtaining undesired sequences.
- Oxidation: transformation of the phosphite triester group into a phosphate triester group by using either I_2 or ^tBuOOH as oxidizing agents.
- Repeat: repeat the cycle n+1 times. The deprotection of the 5'-end of the oligonucleotide can be performed here or after the purification of the oligonucleotide as desired.
- Cleavage: deprotection of the nucleobases and the phosphate group, and cleavage of the oligonucleotide from the resin upon basic treatment with concentrated ammonia solution.

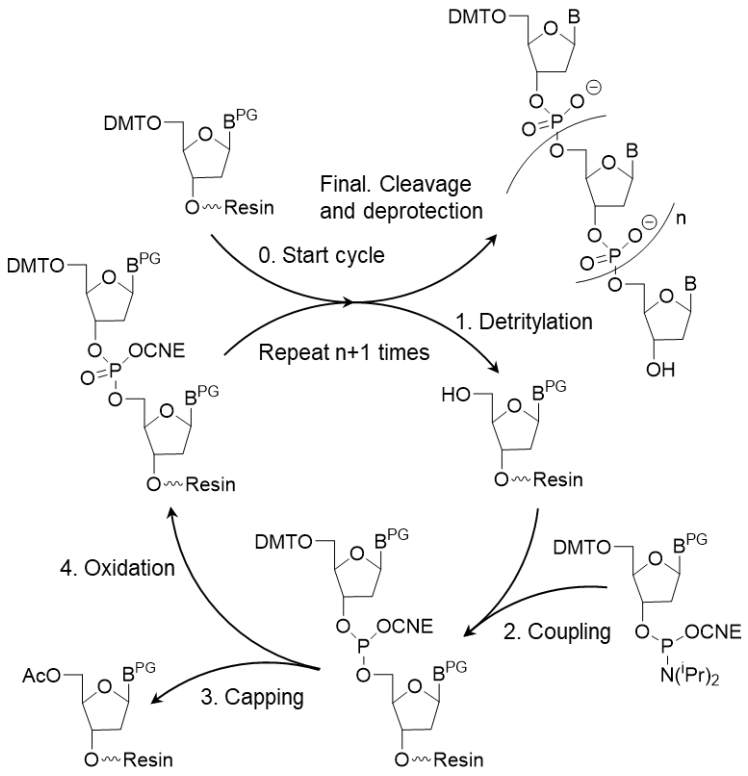


Figure 7-1. Synthetic cycle for linear oligonucleotides

7.1.2. SYNTHESIS OF CYCLIC OLIGONUCLEOTIDES

The cyclic oligonucleotides studied in this thesis were synthesized following a methodology developed in our research group that allows the cyclization step to be performed on the solid support with high selectivity.[253] The overall process, which includes the strategic functionalization of the solid support, is schematically shown in *Figure* 7-2.

The general approach after this methodology includes the use of a TentaGel-NH_2 resin functionalized with the 3'-end nucleotide attached through an o-chlorophenyl linkage. The elongation of the chain is achieved by standard phosphite triester methodology as described in the previous section and the cyclization of the oligonucleotide takes place on-resin by using MSNT as coupling agent. Finally, after the deprotection of the internucleotidic phosphate groups, the cyclic oligonucleotide is selectively cleaved from the solid support through an elimination reaction. For more details, the most outstanding advantages of this method are summarized below:

- The way the resin is functionalized prevents any sequential restrictions since any nucleotide can be incorporated to the solid support. Ideally, pyrimidines are preferred at both ends of the linear oligonucleotide in order to minimize the steric hindrance in the cyclization step.
- Cyclic sequences of up to 30 nucleotides with moderate-to-good yields and a remarkably high purity of the crudes can be obtained.
- The possibility of performing the chain elongation via the phosphite triester method reduces the cost of the synthesis and enables to perform it on an automatic synthesizer.
- The strategic design of the cyclization step on the solid support takes places under dilution conditions, minimizing inter-chain reactions to occur. Moreover, this step can be performed on the automatic synthesizer.
- The specific design of the protection scheme permits to have exclusively deprotected the phosphate group at the 3'-end of the chain allowing a high regioselective cyclization reaction. With the removal of the CNE groups before the final cleavage step, only the phosphate group attached to the solid support remain as phosphate triester, allowing again a selective cleavage of the cyclic oligonucleotide.
- This method tolerates scaling and, thus significant high amounts of cyclic oligonucleotides for different applications such as NMR spectroscopy can be obtained

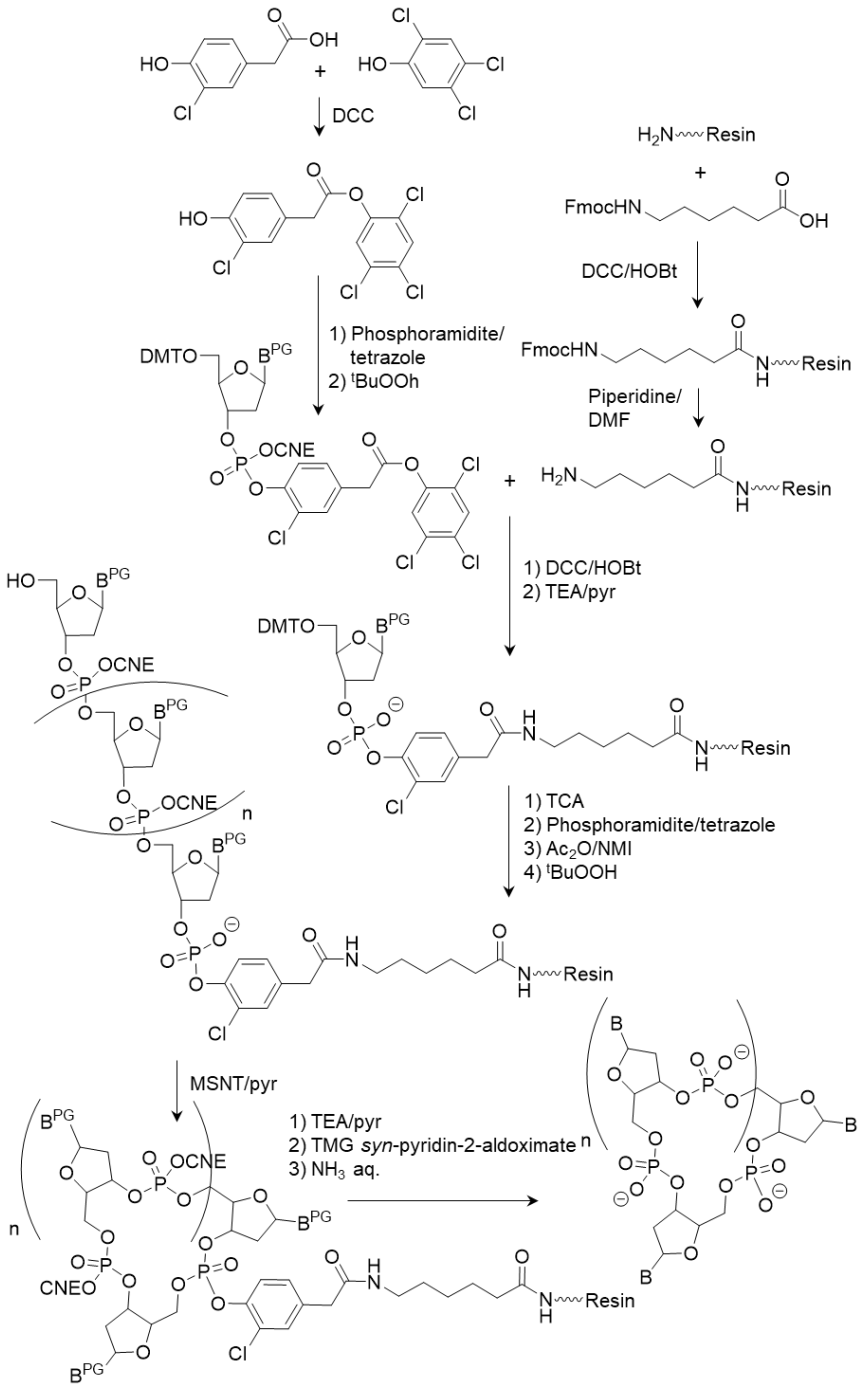


Figure 7-2. Synthetic cycle for cyclic oligonucleotides.

7.2. RESULTS

This section includes all the oligonucleotide syntheses performed during this thesis. The results will include details on the syntheses of the oligonucleotides as well as on the purification and characterization of the sequences. The discussion of these results will be divided by the set of sequences studied on each chapter.

7.2.1. CHAPTER 3: INCORPORATION OF psC IN I-MOTIFS

7.2.1.1. SYNTHESIS

The shorter sequences studied on this chapter (**M**, **MC2**, **M_psC(7)**, **MC2_psC(8)** and **MC2_psC(9)**) were synthesized, purified and characterized in master's degree projects prior to the development of this thesis.[183,184] The syntheses of the psC-containing oligonucleotides derived from the i-motif-forming telomeric sequences were analogously synthesized, following the phosphite triester methodology on a solid support. Contrary to the most common protocol, the following considerations were taken as a precaution due to the use of psC residues:

- UltraMild commercially available phosphoramidites (dC^{Ac}, $dpsC^{dmf}$, $dG^{iPr\text{-}Pac}$), together with dT, were used in order to prevent any harm (deamination or other side reactions) to the psC nucleotides that could potentially occur under the strong basic conditions required for the deprotection of more common phosphoramidites.

- In order to ensure quantitative coupling yields, BTT was used as activator agent and the coupling time for the psC phosphoramidites was enlarged.

The complete synthetic cycle is described in *Section* 8.5 in Methodsas well as the cleavage conditions and the specific details on the quantification, purification, desalting and characterization processes applied to the sequences.

The oligonucleotides were synthesized at 1 µmol scale on the automatic synthesizer starting from a commercial dT-CPG solid support. The previously described considerations due to the incorporation of psC were applied and an RNA-like synthetic cycle was modified in order to perform these syntheses. The final deprotection of the hydroxyl group at the 5'-end of the sequences was performed on the synthesizer. Considering a quantitative coupling yield of 98-99% for each cycle step, the expected yields for the linear synthesis of 21-mer sequences are within the 67-82%. For the first quantification, which was performed after the treatment of the resin with concentrated ammonia at r.t. After the cleavage and deprotection of the oligonucleotides, remarkable crude yields were obtained in a range of 60-67% showing the suitability of the synthesis conditions.

Table 7-1 summarizes the overall results of the syntheses, which discussion is performed below.

Sequence	Crude Yield %	HPLC Purity %	Purified Yield %	HPLC Purity %	Theoretical Mass	MALDI-TOF m/z
HT_psC(1)	60	89	8	>99	6200.1	6198.3
HT_psC(1,7)	67	83	15	>99	6200.1	6198.2
HT_psC(2,8)	62	89	11	>99	6200.1	6197.7

Table 7-1. Results of the syntheses of the sequences studied in Chapter 3.

7.2.1.2. PURIFICATION

After the cleavage of the oligonucleotides, HPLC was used for the analysis of the crude products. Based on the experience learned from the purification of the shorter psC-containing sequences, for which crudes on a reverse-phase C18 column rendered surprisingly wide peaks that strongly hampered their purification, ion-exchange columns were used for the analysis and purification of these sequences. Crudes were analyzed in a 15 to 50% gradient of A in 30 min (A = 1M AcONa 7.5% ACN, B = 7.5% ACN) and purified by using a 25 to 45% gradient of A in 30 min. As shown in *Figure* 7-3, *Figure* 7-4, and *Figure* 7-5, all three oligonucleotides rendered very similar crude profiles of a HPLC purity of 84-89%. The most important drawback of using ion-exchange chromatography for the purification of oligonucleotides was that, given the loss of resolution observed upon increasing the concentration of the samples, the purification had to be done at analytical scale, meaning that a remarkably high number of injections had to be done. Most probably, this great number of injections, together with the high purity criteria followed for the collection of the peaks, led to low yields (8-15%) for the purified oligonucleotides. Nevertheless, as it can be observed in *Figures* 3-3 to 5, the purity of the final products is very remarkable (>99%).

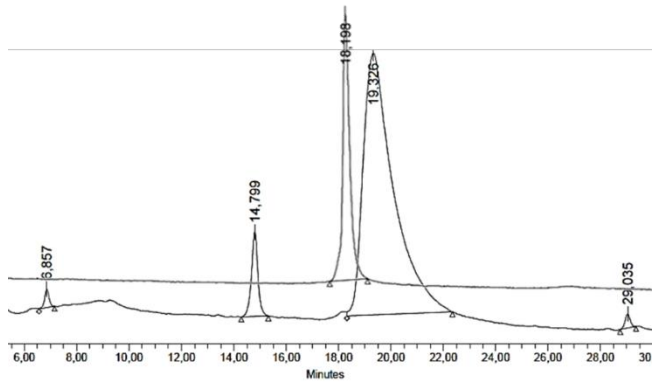


Figure 7-3. Overlapped chromatograms of crude and pure samples of **HT_psC(1)**. 15-50% of A in 30', ion-exchange.

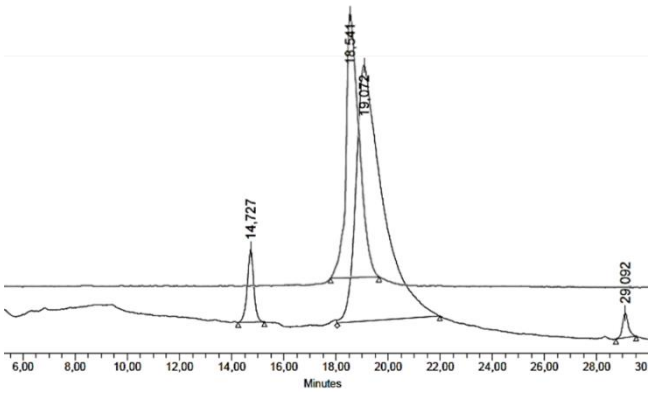


Figure 7-4. Overlapped chromatograms of crude and pure samples of **HT_psC(1,7)**. 15-50% of A in 30', ion-exchange.

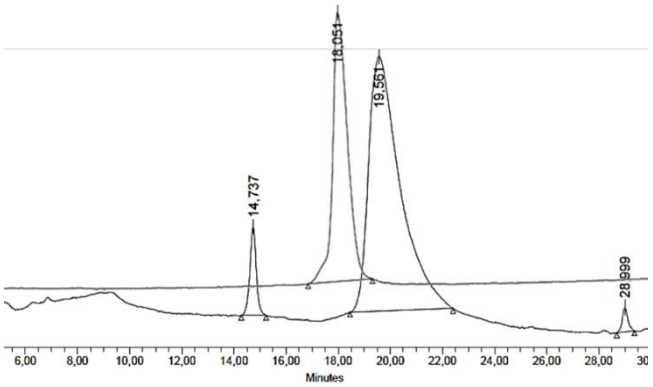


Figure 7-5. Overlapped chromatograms of crude and pure samples of **HT_psC(2,8)**. 15-50% of A in 30', ion-exchange.

7.2.1.3. CHARACTERIZATION

Both pre- and post-purification, the correct identity of the oligonucleotides was checked by MALDI-TOF mass spectroscopy. In *Figures* 7-6 to 8, mass spectra of the purified oligonucleotides are shown. The characterization of these sequences was totally successful as it is reflected in *Table* 7-1. The details on these experiments are explained in *Section* 8.5.1 in Methods.

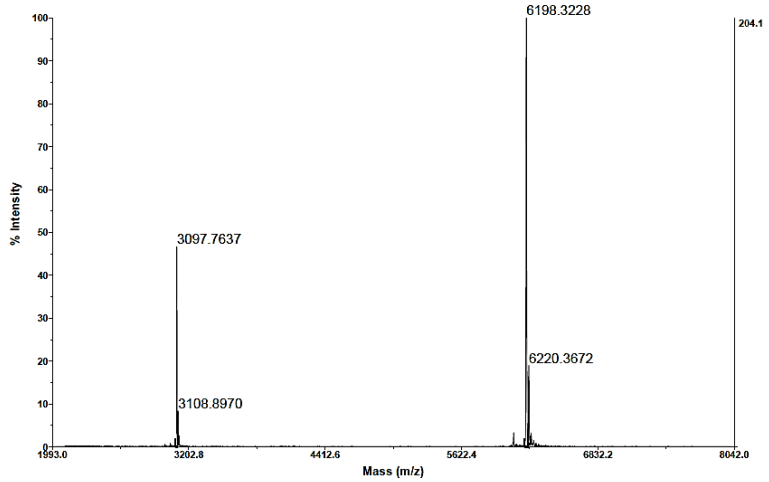


Figure 7-6. MALDI-TOF spectrum of the sequence **HT_psC(1)**.

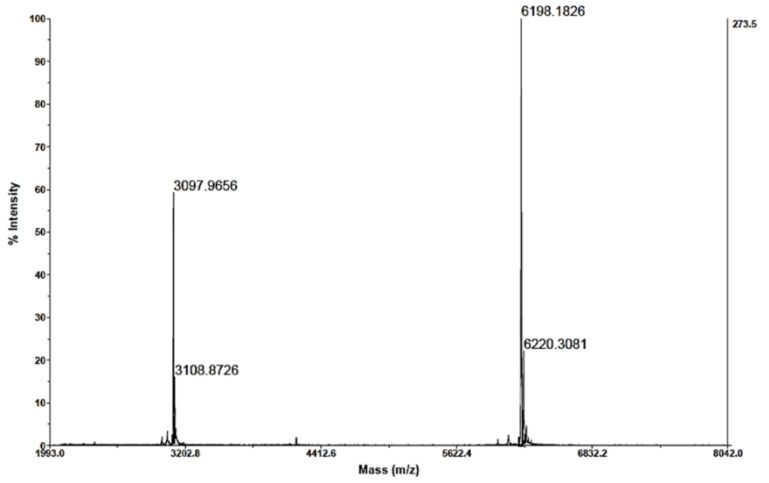


Figure 7-7. MALDI-TOF spectrum of the sequence **HT_psC(1,7)**.

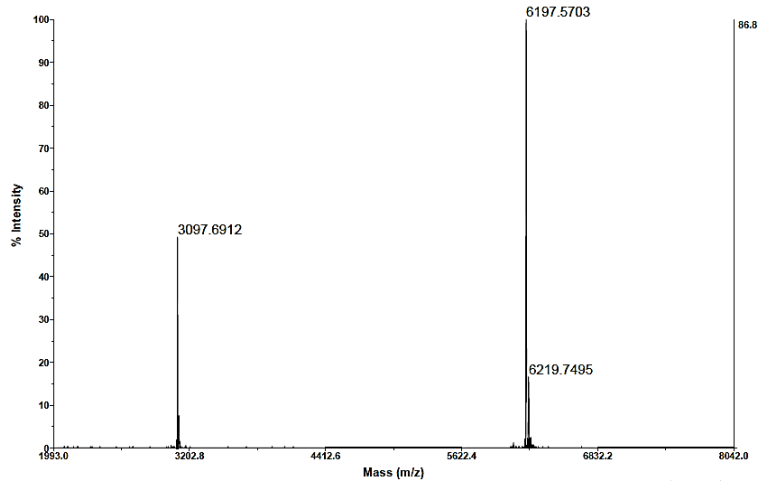


Figure 7-8. MALDI-TOF spectrum of the sequence **HT_psC(2,8)**.

7.2.2. CHAPTER 4: MINOR GROOVE TEDRADS IN I-MOTIFS. PART I: DIMERIC STRUCTURES

Regarding all the oligonucleotides studied in these chapter, **L**, **M** and **N** were purchased from IDT and **cM** was prepared before the beginning of this thesis in our research group.[76] Therefore, the synthetic results regarding this chapter are focused on the cyclic sequences **cL** and **cN**.

7.2.2.1. SYNTHESIS

The cyclic oligonucleotides prepared in this chapter were synthesized following the methodology optimized in our research group that has been described in the Section 1.2 of this chapter. The two syntheses were performed at a 5 μmol scale starting from a cytidinil-derivatized TentaGel resin previously prepared in the research group (f = 105 μmol/g).

7.2.2.2. CHAIN ELONGATION

The syntheses of the linear analogues of **cL** and **cN** (5'-CGTTTCGTTCp-3' and 5'-CGTTCCGTTCp-3', respectively) was carried out following classic phosphite triester methodology. For 10-mer sequences, a global yield for the synthesis of linear oligonucleotides is expected to be 83-91% if the coupling steps are considered quantitative. The report obtained from the automatic synthesizer did not reveal any abnormalities and therefore good yields for this step must be assumed since no real quantification was performed before the cyclization step.

7.2.2.3. CYCLIZATION

The cyclization step for both oligonucleotides was performed on the automatic synthesizer right after the deprotection upon acidic treatment of the 5'-end of the linear sequences. A 0.15 M solution of MSNT in anhydrous pyridine was used to carry out the cyclization of the sequence after checking the melting point of MSNT (136 – 138 ºC) and drying it in a P_2O_5/KOH desiccator. The pyridine and all the material must be strictly anhydrous since hydrolysis of MSNT might occur. Three treatments (4h + 4h + 12h) with a MSNT solution were done.

After the cyclization, different steps for the specific deprotection and cleavage of the cyclic oligonucleotides were done including the removal of the CNE groups, the cleavage of the oligonucleotide from the resin and the deprotection of the nucleobases. Finally, a desalting step must be done in order to remove the residual yet highly concentrated oximate salts used for the selective cleavage of the phosphate triester groups. This desalting step was done by filtering the sample through a Sephadex G-10 column and eluting it with abundant water. The experimental details behind these operations are described in *Section* 8.4.4.1.

After all the steps described above, the quantification of the crudes of the cyclic oligonucleotides rendered low yields that were not completely unexpected for cyclic oligonucleotides (*Table* 7-2). The difficulty of the cyclization step plus the several treatments for the deprotection and cleavage of the oligonucleotides typically leads to low crude yields.

Sequence	Crude Yield %	HPLC Purity %	Purified Yield %	HPLC Purity %	Theoretical Mass	MALDI-TOF m/z
cL	17	50	1.8	99	3047.9	-
cN	20	65	3.4	98	3032.9	3031.2

Table 7-2. Results of the syntheses of the sequences studied in Chapter 4.

7.2.2.4. PURIFICATION

The analysis of the crude via reverse-phase HPLC afforded unusually impure crudes (50-65%) particularly for cyclic oligonucleotides. *Figure* 7-9 and *Figure* 7-10 show the chromatograms of the crude and the purified **cL** and **cN** sequences overlapped. The difference in t_R between the main peaks in the crude and purified chromatograms is attributable to the fact that different chromatographic columns and chromatographic devices were used in each case: the crudes were obtained under semi-preparative conditions while the purified sequences were injected under analytical conditions. In both cases C18 reverse-phase columns were used. The gradient used for the analysis and purification of these oligonucleotides was 7 to 16% of ACN in 25 min. (A = TEAA 0.1 M, B = ACN). After a second purification

required for obtaining reasonably purity values, remarkably low yields were obtained as it is shown in *Table* 7-2. The purity of the final products is remarkably good (98-99%).

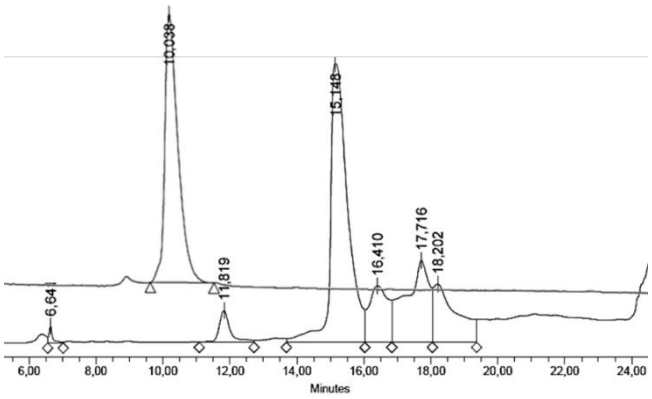


Figure 7-9. Overlapped chromatograms of crude and pure samples of **cL**. 7-16% of B in 25', reverse-phase.

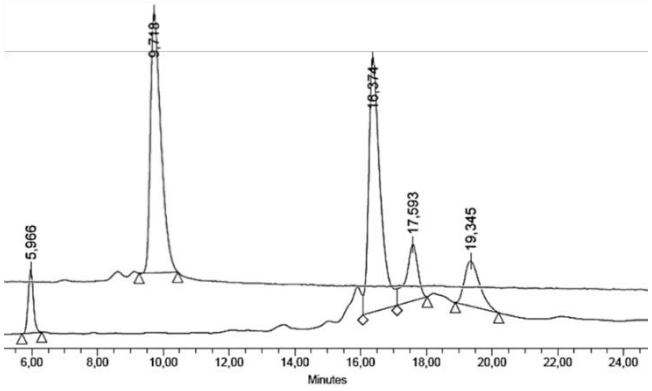


Figure 7-10. Overlapped chromatograms of crude and pure samples of **cN**. 7-16% of B in 25', reverse-phase.

7.2.2.5. CHARACTERIZATION

The characterization of the cyclic sequences **cL** and **cN** was attempted by MALDI-TOF mass spectroscopy. While for **cN** the characterization was successful (*Figure* 7-12), the low oligonucleotide concentration, together with the potential presence of salts in the sample, impeded the correct characterization of **cL** by mass spectroscopy. However, a small peak is observable around 3000 m/z in *Figure* 7-11, probably corresponding to the mass of **cL** (3048 g/mol). However, the characterization of this sequence was unequivocally confirmed from 2D NMR experiments.

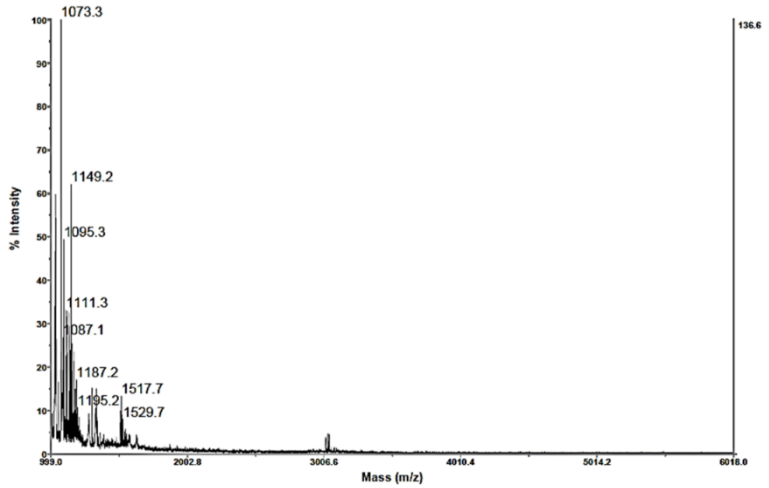


Figure 7-11. MALDI-TOF spectrum of the sequence **cL**.

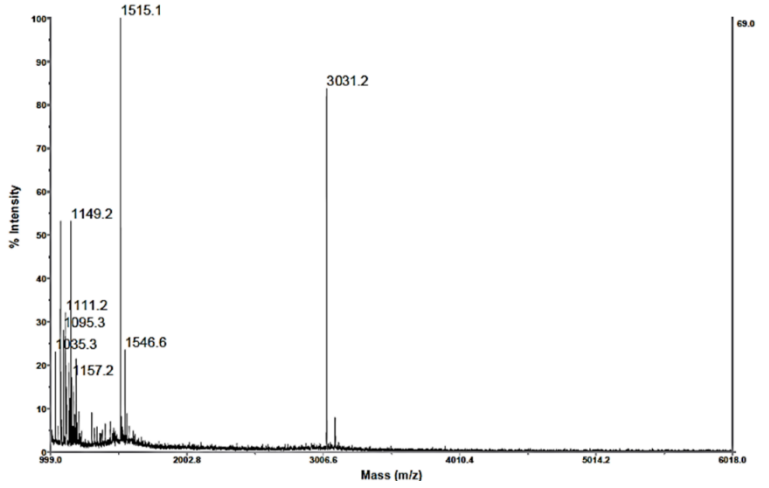


Figure 7-12. MALDI-TOF spectrum of the sequence **cN**.

7.2.3. CHAPTER 5: MINOR GROOVE TEDRADS IN i-MOTIFS. PART II: MONOMERIC STRUCTURES

All the oligonucleotides studied in this chapter were purchased in different amounts from IDT.

7.2.4. CHAPTER 6: FLUORESCENT MINI i-MOTIFS

7.2.4.1. SYNTHESIS

The tC^{o}-containing sequences studied in this chapter were synthesized on an automatic synthesizer following the phosphite triester methodology. Commercially available phosphoramidites were used (dC^{Bz}, dtC^{o}, dG^{iBu} and dT) to perform syntheses at 1 µmol scale. In order to enhance the coupling yield for the tC^{o} step, BTT was used as activator and the coupling time was increased as in the case of psC. The incorporation of tC^{o} residues do not imply any limitation for the application of standard methodology. For a 22-mer oligonucleotide, considering a quantitative yield for each coupling step, the global yield of the linear synthesis should be in the 65-81%. The results for these syntheses are shown in *Table* 7-3. The initial quantification of the crude after the basic treatment for the cleavage and deprotection of the oligonucleotide afforded moderate-to-good yields (52-55%), indicating a proper coupling of the modified nucleotide.

Sequence	Crude Yield %	HPLC Purity %	Purified Yield %	HPLC Purity %	Theoretical Mass	MALDI-TOF m/z
NN4_tC°(2)	55	85	16	97	6699.3	6695.6
NN4_tC°(6)	52	81	12	94	6699.3	6696.3

Table 7-3. Results of the syntheses of the sequences studied in Chapter 5.

7.2.4.2. PURIFICATION

After the synthesis, the oligonucleotides were cleaved in its DMT-on form. This strategy is commonly used for the purification of oligonucleotides since the retention times of the DMT-on sequences is remarkably higher than those of the deprotected deletion sequences, facilitating and speeding quite notably the purification process. The detritylation can be achieved easily afterwards upon acidic treatment (see *Section* 8.5.2.3 in Methods). Crudes were analyzed by reverse-phase HPLC using the following gradient: 0 to 60% of ACN in 40 min (A = TEAA 0.1 M, B = ACN). As it can be observed in *Figure* 7-13 and *Figure* 7-14 for both sequences, two major peaks of similar area appear at ~13 and 20-23 min. These peaks correspond, respectively to the DMT-off and DMT-on sequences. At this point, it was decided that the samples would be completely detritylated in order to optimize its purification on its DMT-off form.

Deprotected sequences were purified by ion-exchange HPLC (A = 1M NaCl 10% ACN, B = 10% ACN) based on the better resolution observed under these conditions. The purification gradient was 25 to 40% of A in 30 min. The purified oligonucleotides where afterwards analyzed again by reverse-phase HPLC obtaining final products 94-97% pure. Medium-to-low final yields were obtained (12-16%) by UV quantification.

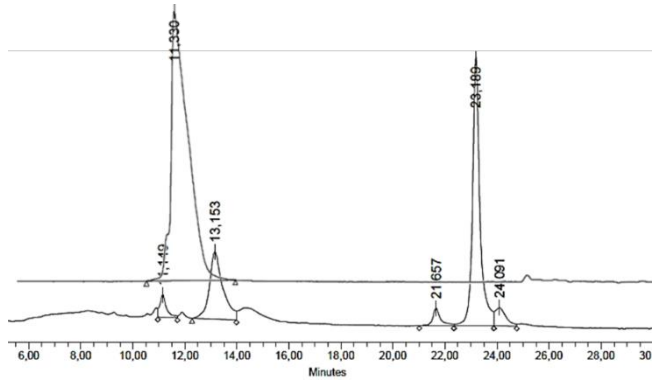


Figure 7-13. Overlapped chromatograms of crude and pure samples of **NN4_tC°(2)**. 0-60% of B in 40', reverse-phase.

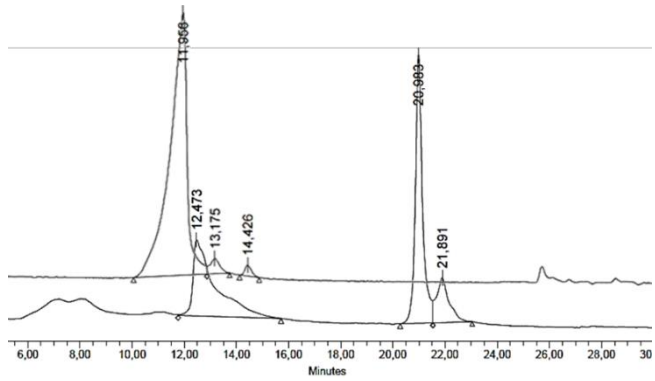


Figure 7-14. Overlapped chromatograms of crude and pure samples of **NN4_tC°(6)**. 0-60% of B in 40', reverse-phase.

7.2.4.3. CHARACTERIZATION

The desalted samples were finally prepared for MALDI-TOF mass spectroscopy in order to assess that purified sequences corresponded to the desired tC°-containing sequences. *Table* 7-3 and *Figures* 16 and 17 show how this characterization was successful.

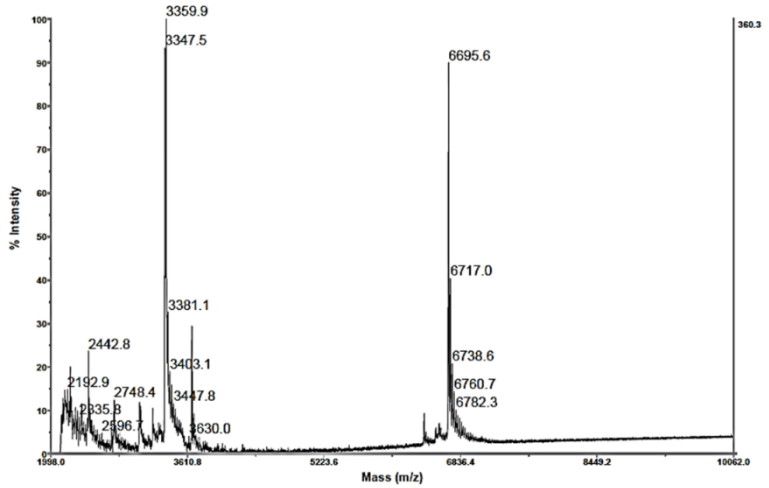


Figure 7-15. MALDI-TOF spectrum of the sequence **NN4_tC^{o}(2)**.

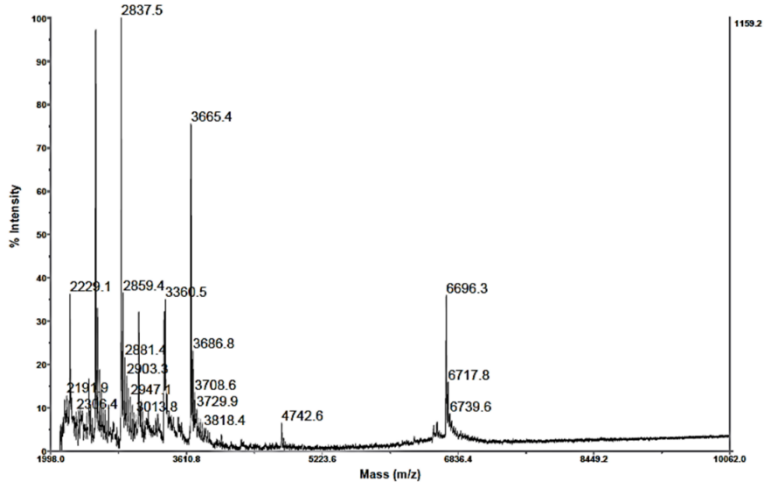


Figure 7-16. MALDI-TOF spectrum of the sequence **NN4_tC^{o}(6)**.

CONCLUSIONS

In view of the results described in this thesis, the following conclusion can be stated:

- The neutral analogue of protonated cytosine, pseudoisocytidine (psC), was incorporated in different positions of dimeric mini i-motifs (**M**-related) and in the telomeric sequence (**HT0)**. The effect of the incorporation of psC depends on its position in the structure, being in most cases destabilizing. Neutral psC:C base pairs stabilize i-motifs at neutral pH, but the stabilization only occurs when psC:C base pairs are located at the ends of intercalated C:C+ stacks. When psC occupies central positions, the resulting i-motifs are only observed at low pH and psC:C+ or psC:psC+ hemiprotonated base pairs are formed, instead of their neutral analogues.

- The solution three-dimensional structure of **MC2_psC(9)** at pH 7 was determined on the basis of experimental constraints obtained from 2D NMR spectroscopy. The structural determination disclosed the formation of a symmetric dimeric i-motif structure adopting a head-to-tail topology held together by two C:C$^+$, two psC:C and two G:G base pairs. The hemiprotonated C:C$^+$ base pairs occupy the core of the structure flanked by two neutral psC:C base pairs. At the ends of the structure, two G:G mismatches are found stacked on the psC:C base pairs.

- Structural and stability data on the incorporation of pseudocytidine in i-motifs suggest that positively charged base pairs in the core of the structure are necessary to stabilize this non-canonical DNA structure. Charge compensation of the high electrostatic repulsion between the sugar-phosphate backbones in the narrow grooves is probably crucial for the stability of the i-motif.

- The effect of different minor groove tetrads in i-motifs was studied in the context of short linear (**M**, **L** and **N**) and cyclic (**cM**, **cL** and **cN**) oligonucleotides, affording dimeric mini i-motifs. The relative orientation of the two sub-units in these dimeric mini i-motifs is strongly conditioned by the nature of the minor groove tetrads capping the C:C$^+$ stack. For symmetric constructions (**M** and **N**), equilibria between the head-to-tail and the head-to head topologies are observed under neutral and acidic conditions. In the particular case of mini i-motifs stabilized by G:C:G:T tetrads, a single species with a head-to-tail topology is observed at neutral pH.

- Unimolecular constructions can be built by connecting the two sub-units of the dimeric mini i-motifs with a poly-T linker. The topology of the resulting structures is all cases head-to-tail, regardless the kind of the minor groove tetrads involved. This linker length affects the stability, but not the formation of the structures.

- Sequences containing **L** repeats were studied in more detail. In particular, the solution structure of **LL3** was determined by restrained molecular dynamics. This structure is stabilized by two C:C$^+$ base pairs capped by two slipped G:C:G:T minor groove tetrads.

- **N**-related linear sequences exhibit a pH-dependent structural transition between two i-motif species. The major species at neutral pH adopts a very similar structure to **LL3**, stabilized with G:C:G:C minor groove tetrads. This acidic form, instead, consists of four hemiprotonated C:C$^+$ base pairs capped by two G:T:G:T minor groove tetrads.

- The consensus sequence d(YCG(XX)YCG(X_n)YCG(XX)YCG) was outlined based on the results obtained for the set of mini i-motif-forming sequences studied. The mapping of this sequence throughout the human genome by bioinformatics analysis revealed a statistical prevalence. The distribution found for these sequences is not random, being much more frequent in in regulatory regions.

- The fluorescent tricyclic phenoxazine analogue tC^o was incorporated in the sequence **NN4** by substituting a single cytosine, either in C:C^+ or in G:C base-pairs. NMR spectroscopy indicates both types of hybridization (tC^o:C^+ and G:tC^o) are compatible with the mini i-motif structure at neutral pH. Interestingly, the fluorescence signal of tC^o suffers a sever quenching when forming an hemiprotonated base pair, compared to the very little quenching that shows upon WC hybridization.

- The solution structure of **NN4_tC^o(2)** was determined by restrained molecular dynamics, based on the NMR experimental data at neutral pH. The formation of tC^o:C^+ base pairs is confirmed, as well as the favorable stacking between tC^o and G:C base pairs from the capping minor groove assemblies. These interactions are responsible for the enhanced stability of this structure against both pH and thermal denaturation.

- It has been probed that tC^o is useful fluorescent probe for the characterization of local environmental transitions in i-motifs. In particular, the transitions between the acidic and neutral species of tC^o-containing **N**-derived sequences were monitored by CD and fluorescence experiments. The transition between the acidic and the neutral species occurs at around pH 6, and the transition towards the unfolded structure, at pH 8.

- Thermal denaturation experiments monitored by fluorescence and CD spectroscopic techniques show similar melting temperatures between **NN4_tC^o(6)** and **NN4**. However, the thermal stability of **NN4_tC^o(2)** is enhanced by almost ten degrees at neutral pH.

- Visualization of mini i-motif structures in cellular media was accomplished by confocal fluorescence microscopy after the successful transfection of HeLa cells with **NN4_tC^o(2)**. Under physiological conditions, the fluorescence emission of tC^o was qualitatively quantified as a 50 % in comparison to a negative control, matching the stability data obtained *in vitro*. The complete quenching of the signal was followed upon acidification, confirming that the folding of the structure translates into a total quenching of the fluorescence emission of the fluorophore tC^o. These results set the basis for the use of tC^o for the detection of i-motif structures in the cell.

METHODS

8. MATERIALS AND METHODS

8.1. REAGENTS, SOLVENTS AND BUFFER SOLUTIONS

8.1.1. GENERAL REAGENTS AND SOLVENTS

AcOH glacial: 99 % solution from Scharlau.

$AcONH_4$: >99 % from Fluka or Sigma-Aldrich.

ACN: HPLC gradient grade (99.99 %) from Fisher Chemical. Anhydrous

ACN: DNA synthesis grade from Applied Biosystems. Add drying traps for synthetic use. Store over CaH_2 and under argon atmosphere to dissolve the phosphoramidites.

Acrylamide: ≥99 % from Fluka.

APS: ≥98 % for molecular biology from Sigma-Aldrich.

Bisacrylamide: 99 % from Sigma-Aldrich.

Concentrated NH_3 solution: 32 % aqueous solution from Scharlau.

CaH_2: 95 % purity from Sigma-Aldrich.

Deionized H_2O (18.2 $M\Omega cm^{-1}$): obtained by Milli-Q plus system from Millipore.

DCM: HPLC gradient grade (99.99 %) from Fisher Chemical. Filter through basic alumina to remove acid traces.

Anhydrous DCM: DNA synthesis grade from Applied Biosystems. Add drying traps for synthetic use. Distill over CaH_2 and store over CaH_2 and under argon atmosphere to dissolve the phosphoramidites.

DMF: 99.8 % solution from Sigma-Aldrich. Store over molecular sieves (4 Å). Let N_2 bubble through the solution to remove volatile amine impurities.

Dioxane: >99 % solution from Sigma-Aldrich. Filter through basic alumina to remove peroxide traces.

EtOH: 99.5 % solution from Panreac.

HCl: 37 % aqueous solution from Scharlau.

I_2: 0.02 M solution in THF/pyr/H_2O from Glen Research or Applied Biosystems.

MeOH: HPLC gradient grade (99.99 %) from Fisher Chemical.

MSNT: 98 % purity from Sigma-Aldrich. Check its purity via melting point (139 ºC) before its use.

NaCl: >99.5 % purity from Fluka or Sigma-Aldrich.

NaOH: lentils from Jescuder.

***Syn*-pyridin-2-aldoxime** : 99+ % purity from Sigma-Aldrich.

Pyridine: 99.8 % anhydrous solution from Sigma Aldrich. Store over CaH_2 and under argon atmosphere.

Sodium cacodylate: ≥98 % from Sigma-Aldrich

TCA: 3 % solution in DCM from Glen Research.

TEA: 99 % solution from Acros Organics or Sigma-Aldrich.

^{t}BuOOH: 5.5 M in decane over molecular sieves (4 Å) from Fluka.

TEMED: 99% from Sigma-Aldrich.

Tetrazole: 0.45 M in ACN from Glen Research. TMG: >99 % purity from Fluka.

8.1.2. BUFFER SOLUTIONS AND HPLC ELUENTS

- $AcONH_4$ 1 M, 7.5% ACN, pH 7:

For a 1 L solution dissolve 77.1 g of $AcONH_4$ (>99 % purity from Fluka or Sigma-Aldrich) in water and adjust to pH 7 with NH_3 or AcOH. Add 75 ml of HPLC-grade ACN. Level up to 1 L with water. The solution is filtered (45 µm pore size) for HPLC use. Store at 4 ºC.

- NaCl 1M, 10% ACN:

For a 500 mL solution dissolve 29.2 g of high purity NaCl (>99 % from Fluka or Sigma-Aldrich) in water and check the pH. Level up to 500 mL with water. The solution is filtered (45 µm pore size) for HPLC use. Store at 4 ºC.

- TEAA 2 M, pH 7:

For a 1 L solution add gently, while stirring, 280 mL of TEA and 116 mL of AcOH to 500 mL of water cooled in an ice bath. Adjust to pH 7 with TEA or AcOH and level up to 1 L with water. The solution is filtered (45 µm pore size) for HPLC use. Store at 4 ºC.

- NaPi 0.2 M, pH 7:

For a 50 mL solution dissolve 0.53 g of $NaH_2PO_4 \cdot 2H_2O$ and 1.10 g of $Na_2HPO_4 \cdot 2H_2O$ (>99 % purity from Fluka). Check the pH, adjust if necessary and level up to 50 mL. Store at 4 ºC.

- Cacodylate 25 mM, 100 mM NaCl, pH 5.5:

For a 50 mL solution dissolve 270 mg of sodium cacodylate and 290 mg of NaCl in water. Check the pH and adjust it with HCl to 5.5. Level up to 50 mL with water and store at 4 ºC.

8.2. INSTRUMENTATION

8.2.1. SPECTROSCOPIC TECHNIQUES

- NMR

All the oligonucleotide NMR spectra were recorded on Bruker AV-600 and Bruker AV-800 US2 devices from the Spectroscopy and Molecular Structure Department of the "Instituto de Química-Física Rocasolano", which belongs to the CSIC in Madrid.

- UV spectroscopy

Quantification UV-vis spectra were recorded on a SECOMAM UVIKON XS device from the Organic Chemistry Department or on a Jasco V-730 spectrophotometer from our research group. The melting curves were obtained exclusively from the Jasco V-730 spectrophotometer, which includes a thermostatically regulated six-positions carrier coupled to Peltier Jasco PAC-743. Quartz cuvettes of 1 cm path length from Hellma were used.

- Circular dichroism

Circular dichroism spectra and CD melting experiments were recorded on a Jasco J-815 device coupled to a Peltier Jasco CDF-4265 accessory from the CCiTUB. Quartz cuvettes of 1 cm path length from Hellma were used.

- Fluorescence

Fluorescence spectra were recorded on a Photon Technologies International spectrofluorometer equipped with a xenon lamp and a four-position carrousel thermoregulated by an external PEG/H_2O bath. Quartz cuvettes of 1 cm path, special for fluorescence applications, from Hellma were used.

8.2.2. MASS SPECTROSCOPY

MALDI-TOF mass spectra were recorded on an ABSciex 4800 plus device from the mass spectrometry department from the CCiTUB, located in the Faculty of Chemistry.

8.2.3. CHROMATOGRAPHIC TECHNIQUES

- HPLC

Oligonucleotide analyses and purifications were carried out in Waters devices: Separations Module 2695 and PDA Detector 2996 both for analytical and low scale purification use. Controller 600, Autosampler 717plus and Dual λ Absorbance Detector 2487 for higher scale purification.

For analytical conditions a reverse phase C18 Jupiter column from Phenomenex was used (size: 250 x 4.6 mm, particle size: 10 µm). Flow rate: 1 mL/min.

For semi-preparative scale purifications, a reverse phase C18 Jupiter column from Phenomenex was used (size: 250 x 10 mm, particle size: 10 µm). Flow rate: 3 mL/min.

For ion exchange analyses and purifications, NucleoPac PA-100 column from Dionex was used (size: 250 x 4 mm, composition: 13 µm-diameter nonporous substrate agglomerated with a 100 nm MicroBead alkyl quaternary ammonium functionalized latex). Flow rate: 1.5 mL/min.

Eluents used for reverse phase columns were TEAA 0.1 M (line A) and ACN (line B).

Eluents used for the ion exchange column were $AcONH_4$ 1M pH 7 / ACN 7.5% (line A) and H_2O / ACN 7.5% (line B) or NaCl 1M / ACN 10% (line A) and H_2O / ACN 10 % (line B).

8.2.4. POLYACRYLAMIDE GEL ELECTROPHORESIS

20% polyacrylamide gels (acrylamide/bisacrylamide 19:1) of 0.75 mm were prepared in a Mini-PROTEAN Tetra Cell system from Bio-Rad. The gels were run with an Electrophoresis Power Supply EPS 3501 XL.

8.3. OTHER INSTRUMENTATION

The lyophilization of the samples was carried out on machines from Labconco (model Freezone 6) and Virtis (model Freezemobile 12EL). The pH measurements were performed using a Crison micropH Basic 20 device. Centrifugations were carried out on a 5430R device from Eppendorf. A Thermovar device was used for the determination of melting points. A SpeedVac device from ThermoFisher was used for the evaporation of ammonia after oligonucleotide cleavage.

8.4. SOLID PHASE OLIGONUCLEOTIDE SYNTHESIS

8.4.1. REAGENTS, SOLVENTS AND SPECIFIC INSTRUMENTATION

All the sequences where synthesized by an automatic synthesizer 3400 from Applied Biosystems. Different synthetic cycles were carried out depending on the sequence, the synthesis scale and the polymeric support.

The polymeric supports used for the syntheses were the following:

- PS-PEG copolymer TentaGel® N NH_2 from RAPP POLYMERE for cyclic oligonucleotides synthesis. (f = 240 µmol/g).

- Nucleoside functionalized CPG support (5'-DMT-2'-deoxy-3'-succinoyl-lcaa-CPG) from Glen Research for linear oligonucleotides synthesis. (f = 40-50 µmol/g).

The 3'-(2-CNE)-N,N-$(iPr)_2$-phosphoramidites of 5'-O-(4,4'-DMT)-deoxynucleosides used were from Glen Research or Link Technologies. UltraMild (DMT-dC^{Ac}-CNE and DMT-$dG^{iPr\text{-}Pac}$-CNE) protected phosphoramidites were used for the synthesis of psC-containing oligonucleotides. DMT-dC^{Bz}-CNE, DMT-dG^{iBu}-CNE and DMT-dT-CNE) phosphoramidites were used for the rest of the syntheses. DMT-$dpsC^{dmf}$-CNE from Berry&Associates and DMT-dtC^{o}-CNE from Glen Research phosphoramidites were used for the synthesis of modified sequences.

The reagents used for the non-modified sequences were the following and are from Glen Research or Applied Biosystems:

- Activator: Tetrazole 0.45 M in ACN.

- Capping solutions: Cap Mix A: Ac_2O/THF/pyr. Cap Mix B: 16 % NMI in THF.

- Detritylating solution: TCA 3% in DCM.

- Oxidant: I2 0.02 M solution in THF/pyr/H2O.

For psC-containing sequences, RNA synthesis reagents (BTT and ^{t}BuOOH) were used. In the case of tC^{o}-containing sequences, BTT (crystalline, Link Technologies) was also used as activator agent.

The phosphoramidites, previously dried for 12 hours in a vacuum desiccator with P_2O_5 and KOH, were dissolved in anhydrous ACN or in anhydrous DCM for the synthesis using CPG and TentaGel® as polymeric supports, respectively. The resins were weighted directly on the Teflon reactors that were further sealed with aluminum rings. The phosphoramidites were always manipulated under argon atmosphere. The argon used was 5.0 quality with less than 3 ppm/v of water. All the glass material was dried at 110 ºC for 12 hours and vacuum dried at the desiccator.

8.4.2. OLIGONUCLEOTIDES QUANTIFICATION BY UV SPECTROSCOPY

The parameter used to determine the quantity of oligonucleotide was optic density. A unit of optic density (OD) defines the quantity of oligonucleotide for 1 mL of sample that exhibits an absorbance of 1 in a 1 cm of path length cuvette. According to this, and picking the maximum absorbance around 260 nm, the oligonucleotides were quantified as it follows: $OD_{max} = Abs_{max} x\ V\ (ml)$.

In order to obtain the molar equivalency from this density units ($OD_{max} = \varepsilon\ (mL \cdot \mu mol^{-1} \cdot cm^{-1})\ xn\ (\mu mol)$), molar extinction coefficients for each oligonucleotide were calculated by the nearest-neighbor method using the expressions and parameters described by Owczarzy.[254]

The molar extinction coefficients of the nucleotides synthesized for this work are presented below (*Tables* 8-1 to 4). The molar extinction coefficients for the psC-containing sequences were calculated considering that cytidine and pseudoisocytidine have the same molar extinction coefficient. The molar extinction coefficients of the tC^{o}-containing sequences were calculated as reported by Sandin *et. al.*[255]

SEQUENCE	ε (ml·µmol-1·cm-1)
M, M_psC(7)	78.1
MC2, MC2_psC(8), MC2_psC(9)	92.5
HT0, HT_psC(1), HT_psC(1,7), HT_psC(2,8)	185.9

Table 8-1. Molar extinction coefficients of all the sequences studied on Chapter 3.

SEQUENCE	ε (ml·µmol-1·cm-1)	SEQUENCE	ε (ml·µmol-1·cm-1)
M	78.1	cM	87.0
L	78.8	cL	86.7
N	75.6	cN	85.6
K	82.3		

Table 8-2. Molar extinction coefficients of all the sequences studied on Chapter 4.

SEQUENCE	ε (ml·µmol-1·cm-1)	SEQUENCE	ε (ml·µmol-1·cm-1)
LL1	161.9	LL7	210.5
LL2	170.0	LL3rep	379.9
LL3	178.1	LL3long	404.2
LL4	186.2	LL4rep	404.2
LL5	194.3	MM4	188.0
LL6	202.4	NN4	183.7

Table 8-3. Molar extinction coefficients of all the sequences studied on Chapter 5.

SEQUENCE	ε (ml·µmol-1·cm-1)
NN4	183.7
NN4_tC^{o}(2)	186.1
NN4_tC^{o}(6)	186.1

Table 8-4. Molar extinction coefficients of all the sequences studied on Chapter 6.

8.4.3. OLIGONUCLEOTIDES PURIFICATION

The purifications of the synthesized oligonucleotides were carried out exclusively via HPLC. Both reverse phase and ion exchange columns were used. Further purification issues are discussed individually for each synthesis in Chapter 7.

8.4.4. OLIGONUCLEOTIDES DESALTING

8.4.4.1. DESALTING BY MOLECULAR EXCLUSION

Some oligonucleotides were desalted by molecular filtration through Pharmacia's Sephadex G-10 columns (80x2 cm) in order to remove the excess of oximate salts originated from the cleavage of cyclic oligonucleotides. The assembled system includes a Pharmacia LKB P-1 peristaltic pump, an Endress+Hauser RSG 30 UV-vis detector and a Pharmacia Biotech REC 101 recorder.

8.4.4.2. DESALTING BY CATIONIC EXCHANGE

In order to substitute the TEA cations from samples after the HPLC purification by sodium cation a Dowex 50Wx4 (200-400 mesh, 4.8 meq. Na^{+}/g dry resin) was used. The resin is placed on a polypropylene syringe equipped with a porous polyethylene filter and the following protocol is followed:

1. Wash the resin with water.
2. Pass through the column 200 mL of HCl 10 %.
3. Wash the resin with water until reaching initial pH.
4. Pass through the column 200 mL of NaOH1 M.
5. Repeat step 3.
6. Charge the oligonucleotide dissolved in 1-2 mL H2O.
7. Elute with water. Collect fractions of 2-3 mL. Check product presence by UV spectroscopy.

8.4.4.3. DESALTING BY CENTRIFUGAL DEVICES

Amicon® Ultra devices from Millipore were used for desalting samples for NMR and spectroscopic studies. Dissolve the samples in 0.5 mL of water and charge the membrane filter with the solution. Centrifuge for 15 min at 14000 rcf. Flip the membrane filter on a new Eppendorf and centrifuge for 2 min at 1000 rcf to recover the desalted oligonucleotide. Refill the membrane filter with the desired buffer solution or water and repeat the procedure for 5-6 times. This method is only valid for samples with a molecular weight higher than 3000 g/mol.

8.4.5. MALDI-TOF MS CHARACTERIZATION

The samples are prepared by mixing 1 µl of oligonucleotide solution (100-500 µM) with 1 µl of ammonium citrate solution and 1 µl of THAP as matrix. The used solutions are the following:

- Ammonium citrate 98 % from Sigma-Aldrich. 50 mg/ml solution in H2O.

- 2,4,6-trihidroxyacetophenone (THAP) 98 % from Sigma-Aldrich. 10 mg/ml solution in H2O/ACN 1:1.

The spectra were recorded on an ABSciex 4800 plus device. Negative mode was used to the nature of the sample. Reflector mode was used for oligonucleotides with a mass up to ~3500 g/mol. Linear mode was used for higher mass oligonucleotides or in difficult observation situations. Linear mode presents higher sensitivity but lower accuracy.

8.5. LINEAR OLIGONUCLEOTIDE SYNTHESIS

All the syntheses carried out in this thesis have been obtained according the phosphite triester methodology described in the previous chapter. The phosphoramidites were prepared and manipulated as described above. The autodilution mode of the synthesizer was used adjusting the parameters for each synthesis in order to obtain 0.1M phosphoramidite solutions. Commercial CPG functionalized resins were used. *Table* 8-5 summarizes the synthetic cycle designed for the synthesis of non-modified linear sequences at 1µmol scale. Based on this synthetic cycle, the modifications applied depending on the nature of the sequences will be highlighted individually in the following sections.

STEP	OPERATION	REAGENT/SOLVENT	DELIVERY (s)	SLEEP (s)
1	Wash	ACN	10	0
2	Detritylation	TCA 3% in DCM	110	0
3	Wash	ACN	15	0
4	Dry	Argon	10	0
5	Coupling	Phosphoramidite + BTT in ACN	5	30
6	Wash	ACN	15	0
7	Capping	Ac_2O/NMI	12	6
8	Wash	ACN	15	0
9	Oxidation	I_2 0.1 M in H_2O/pyr/THF	12	20
10	Wash	ACN	15	0

Table 8-5. Step by step description of the standard synthetic cycle for linear oligonucleotides.

8.5.1. psC-CONTAINING SEQUENCES

8.5.1.1. CHAIN ELONGATION

RNA-synthesis-like methodology was used. BTT 0.3 M in anhydrous ACN was used as activator agent. ^{t}BuOOH 1 M in DCM was used as oxidant agent. The previously weighted UltraMild phosphoramidites (dC^{Ac}, $dpsC^{dmf}$, $dG^{iPr\text{-}Pac}$, dT) were dissolved in anhydrous ACN using the autodilution mode of the synthesizer. The synthetic cycle was modified in order to enlarge the coupling time (1200 s) for the psC residue. The final detritylation of the 5'-end of the sequences was performed on the solid support on the automatic synthesizer.

8.5.1.2. CLEAVAGE AND DEPROTECTION

After the chain elongation, the resins are transferred from the reactor to a cleavage vial. Concentrated NH_3 solution was added (1-2 mL) and the cleavage vial was stored at r.t. for 12h to accomplish the cleavage and the deprotection of the nucleobases. After the 12h, the resin is removed from the oligonucleotide solution by adding H_2O and filtering the solution with a PTFE filter coupled to a syringe. The ammonia excess is removed by evaporation and the samples are lyophilized to dryness.

8.5.1.3. PURIFICATION AND DESALTING

The psC-containing sequences were purified by ion-exchange chromatography as described in the Syntheses *Section* 7.2.1.2Once purified, the samples were desalted using Amicon® Ultra centrifugal devices.

8.5.2. tC^{o}-CONTAINING SEQUENCES

8.5.2.1. CHAIN ELONGATION

The synthesis of the tC^{o}-containing sequences was achieved following the synthetic cycle described in *Table* 8-5. BTT 0.3 M in anhydrous ACN was used as activator agent. The previously weighted phosphoramidites (dC^{Bz}, dtC^{o}, dG^{iBu}, dT) were dissolved in anhydrous ACN using the autodilution mode of the synthesizer. The coupling time for the tC^{o} residues was enlarged compared to the other nucleobases (900 s). The syntheses ended without the final detritylation step in order to optimize their purification.

8.5.2.2. CLEAVAGE

After the chain elongation, the resins are transferred from the reactor to a cleavage vial. Concentrated NH_3 solution was added (1-2 ml) and the cleavage vial was stored at 55 ºC for 12h to accomplish the cleavage and the deprotection of the nucleobases. After the 12h, the samples are cooled down at r.t. and the resin is removed from the oligonucleotide solution by adding H_2O and filtering the solution with a PTFE filter coupled to a syringe. The ammonia excess is removed by evaporation and the samples are lyophilized to dryness.

8.5.2.3. DETRITYLATION

After the failed attempt of performing the purification of these sequences DMT-on, the sequences were detrityilated before continuing with their purification. The acidic deprotection with AcOH was carried out by treating the dried samples with 1.5 mL solution of AcOH 80 % (v/v) for 25 min at 0 ºC. The samples are afterwards diluted to a total volume of 15 mL with water. In order to remove the DMT from the solution, three extractions with ether were performed (3x10 mL). Finally, the samples were lyophilized.

8.5.2.4. PURIFICATION AND DESALTING

The tC^{o}-containing sequences were purified by ion-exchange chromatography as described in the Syntheses *Section* 7.2.2.4. Once purified, the samples were desalted using Amicon® Ultra centrifugal devices

8.6. CYCLIC OLIGONUCLEOTIDE SYNTHESIS

The synthetic approach that was used for the cyclic oligonucleotides prepared in this thesis follows the methodology described in our research group that was deeply explained in the synthetic chapter of this manuscript.[253] The following sections present experimental detail on the different steps of these syntheses.

8.6.1. CHAIN ELONGATION

A cytidinil-derivatized TentaGel® resin, with unprotected 3'-phosphates groups,previously prepared in the group with a functionalization of 105 μmol/g was employed. The resin was weighted on a Teflon reactor and sealed with aluminum rings. The chain elongation was carried out at 5 μmol scale. The commercially available cyanoethyl phosphoramidite derivatives of the nucleosides (dC^{Bz}, dG^{iBu} and dT) were manipulated as described above and dissolved in anhydrous DCM under argon atmosphere. The DCM was previously filtered through basic alumina in order to remove any acidic traces that would potentially detritylate the phosphoramidites. Commercial tetrazole solution was used as activator agent and ^{t}BuOOH was used as oxidizing agent. The synthesis cycle that was used for the elongation of the linear chain is summarized in *Table* 8-6. As it can be observed, the coupling times for all the residues are significantly high in order to ensure quantitative coupling yields.

STEP	OPERATION	REAGENT/SOLVENT	DELIVERY (s)	SLEEP (s)
1	Wash	ACN	20	0
2	Detritylation	TCA 3% in DCM	120	0
3	Wash	ACN/DMF/DCM	150/40/60	0
4	Dry	Argon	50	0
5	Coupling	Phosphoramidite + tetrazole in DCM	20	900
6	Wash	ACN	30	0
7	Capping	Ac_2O/NMI	30	120
8	Wash	ACN	50	0
9	Oxidation	^{t}BuOOH 1 M in DCM	40	90
10	Wash	DCM/DMF/ACN	40/50/50	0

Table 8-6. Step by step description of the synthetic cycle for the chain elongation of cyclic oligonucleotides.

8.6.2. CYCLIZATION

After deprotection of 5'-OH group at the end of the linear chain, the cyclization reaction was performed on the same reactor in the automatic synthesizer under argon atmosphere. The cyclization reagent was a 0.15 M solution of MSNT in anhydrous pyridine. After checking its melting point (136 – 138 ºC), the MSNT was dried and stored in a P_2O_5/KOH desiccator. The pyridine and all the material must be strictly anhydrous. Three treatments (4h + 4h + 12h) with MSNT were done. The procedure is summarized in *Table* 8-7. The MSNT must be dissolved immediately before its use and renovated for each treatment.

STEP	OPERATION	REAGENT/SOLVENT	DELIVERY (s)	SLEEP (h)
1	Wash	ACN	40	0
2	Dry	Argon	10	0
3	Wash	Pyridine	30	0
4	Dry	Argon	15	0
5	Cyclization	MSNT 0.15 M in anh. pyr	30	4 + 4 + 12

Table 8-7. Cyclization cycle.

8.6.3. CLEAVAGE AND DEPROTECTION

- Deprotection of the diester phosphates:

After the cyclization step the resin was transferred to a syringe equipped with a polypropylene filter and washed with DCM. To accomplish CNE removal, the resin was treated with TEA/pyr 1:1 (3x1h).

- Cleavage of the triester phosphate:

In order to cleave selectively the *o*-chlorophenyl phosphate triester linkage between the cyclic oligonucleotides and the solid support, a 0.2 M solution of *syn*-pyridin-2-aldoximate of TMG in H2O/dioxane 1:1 was used. Three treatments for 4 + 4 + 12 hours were done. The resin was properly washed between treatments with H_2O/dioxane 1:1, and all the filtrates, that contained the cyclic product in solution, were collected.

- Deprotection of the nucleobases:

All the collected fractions from the cleavage were combined and treated with twice their volume of concentrated NH_3 at 55 ºC during 12 h, subsequently transferred to a 50 mL round-bottom flask and evaporated to dryness.

8.6.4. PURIFICATION AND DESALTING

Prior to purification, the samples were passed through a Sephadex G-10 column in order to remove the excess of oximate salts. The samples are charged in the minimum volume of water/buffer and eluted with abundant water. The purification of the cyclic oligonucleotides was accomplished by reverse-phase HPLC as it was described in *Section* 7.2.4.2. The final desalting of the pure products was performed with Amicon® Ultra centrifugal devices.

8.7. BIOPHYSICAL EXPERIMENTS

8.7.1. SAMPLE PREPARATION

The procedure for the preparation of the different oligonucleotide samples studied in this thesis by UV, DC and/or Fluorescence spectroscopy is, in general terms, independent of the technique used for the analysis. The samples were prepared from more concentrated solutions of both oligonucleotide and buffer solutions considering the final volume desired depending on the cuvette. Generally, the desired amount of pure and desalted oligonucleotide was lyophilized and then dissolved with the desired final volume of the buffer solution at the right concentration. The renaturation of the samples was performed in a sand bath heated at 90 ºC and tempered smoothly at r.t. After this procedure the samples were stored in the fridge (4 ºC) for at least 8 - 12 h.

8.7.2. SPECTRA ACQUISITION

The CD spectra at different temperatures were recorded on a Jasco J-815 device coupled to a Peltier Jasco CDF-4265 accessory. Before starting the acquisition, the samples were allowed to stabilize for 5 min inside the instrument. A N_2 flow rate between 5 and 10 ml/min is maintained during the working time. The spectra were recorded at different temperatures, scanning from 320 to 220 nm. Each spectrum is the result of three accumulations. The blank correction was done after the recording of the spectra by deducting the ellipticity at 320 nm after checking that the buffer does not exhibit any ellipticity.

The fluorescence excitation and emission spectra were analogously recorded on a Photon Technologies International spectrofluorometer equipped with a xenon lamp and a four-position carrousel thermoregulated by an external PEG/H_2O bath. The spectra were recorded at different temperatures, allowing the sample to stabilize for 5 min before starting the acquisition. Excitation spectra were recorded from 230 to 420 nm, detecting at 450 nm. The emission spectra were recorded from 380 to 580 nm, prior excitation at 370 nm.

8.7.3. MELTING EXPERIMENTS

The melting curves monitored either by UV or CD spectroscopy were programmed analogously. The renatured samples taken from the fridge were allowed to stabilize at the starting temperature (commonly 5 ºC) for 10 min before starting the temperature gradient. The variation of the ellipticity or the absorbance was monitored at a fixed wavelength (commonly 260 - 270 nm) and a 0.5 ºC/min temperature rate was applied until denaturation (65 - 85 ºC) by the Peltier accessory. If desired, the software of both devices includes the possibility of programming pauses during the temperature ramp (and then resume it) and record a spectrum in a range of desired wavelengths (320 - 220 nm in our case) at any given temperature in order to follow potential variations at other wavelength values.

8.7.4. pH-TITRATION EXPERIMENTS

Analogously to the description of the experiments presented in Section 4.2, the monitorization of the CD and fluorescence spectra was performed upon varying the pH of the samples at 5 ºC. For these experiments, the pH of the samples was adjusted to acidic values (3.5 - 4) before renaturing them. Once renatured, the samples were taken from the fridge and allowed to stabilize at the cell holder for 10 min before starting the experiment. Then, spectra analogous to the ones described in Section 4.2 were recorded at different pH values until total denaturation of the samples (pH 9 – 10) was observed. These pH variations were accomplished by addition of low volume aliquots (0.5 – 2 µL) of a concentrated solution of NaOH. The pH of the sample was checked externally by a Crison pHmeter before and after the acquisition of the spectra. Between acquisitions the samples were allowed to stabilize for 5 min.

8.7.5. DATA TREATMENT

8.7.5.1. pH_T AND T_m VALUES

The normalized data from the melting experiments and the raw data from the pH-titration experiments render sigmoidal plots when represented against temperature or pH. From these experimental curves, as it was introduced in Fundamentals, very interesting parameters such as T_m or pH_T can be extracted.

Using OriginPro 8 tools, the experimental data was adjusted to a sigmoidal fit using Boltzmann models, hereafter represented for CD-monitored pH-titration experiments as example:

$$Y = \varepsilon_{unfolded} + \frac{\left(\varepsilon_{folded} - \varepsilon_{unfolded}\right)}{1 + e^{\left(\frac{pH_T - pH}{dx}\right)}}$$

Associated error in T_m values determination was estimated in 0.5 °C. Associated error in pH_T determination was estimated in 0.1 pH units.

8.7.5.2. THERMODYNAMIC PARAMETERS

The thermodynamic parameters presented in Chapter 5 were calculated by the theoretical approach presented in Fundamentals. Summarized, this approach applies a Van't Hoff plot of the experimental data obtained from melting experiments in order to calculate thermodynamic parameters:

$$K = -\frac{1}{T} \cdot \frac{\Delta H°}{R} + \frac{\Delta S°}{R}$$

K values arise from the following equation, which includes the molar fraction (x) of structured species at any given temperature which is calculated from the variation of a spectroscopic magnitude as a function of the temperature. For structural transitions involving exclusively one molecule, the equation for calculating the *K* values is the following:

$$K = \frac{x}{(1-x)} \qquad x = \frac{A - A_{unfolded}}{A_{folded} - A_{unfolded}}$$

Determination of thermodynamic data is ultimately achieved from a Van't Hoff plot, which fits the representation of ln*K* against 1/T to a linear model. Usually, given that the representation of ln*K* against the temperature renders not completely linear traces, that is why the chosen x values comprise a range of central values that afford more valid *K* values. The range chosen for these calculations in this thesis were 0.15<x<0.85 in order to obtain trustable results. The plots and the linear fits, together with the obtained linear equations are shown in *Figure* 8-1 and were obtained from OriginPro 8 analysis applications.

Deviations from parameters have been calculated by applying the error propagation method. For calculating associated variances, the following equations have been applied:

$$(\sigma_{\Delta H^0})^2 = \sigma_m^2 \cdot (-R)^2$$

$$(\sigma_{\Delta S^0})^2 = \sigma_b^2 \cdot (R)^2$$

$$(\sigma_{\Delta G^0})^2 = \sigma_b^2 \cdot (-TR)^2 + \sigma_m^2 \cdot (-R)^2 + 2\sigma_{bm}^2 \cdot (-TR) \cdot (-R) \qquad \left(\sigma_{bm}^2 = \sqrt{\sigma_b^2 \cdot \sigma_m^2}\right)$$

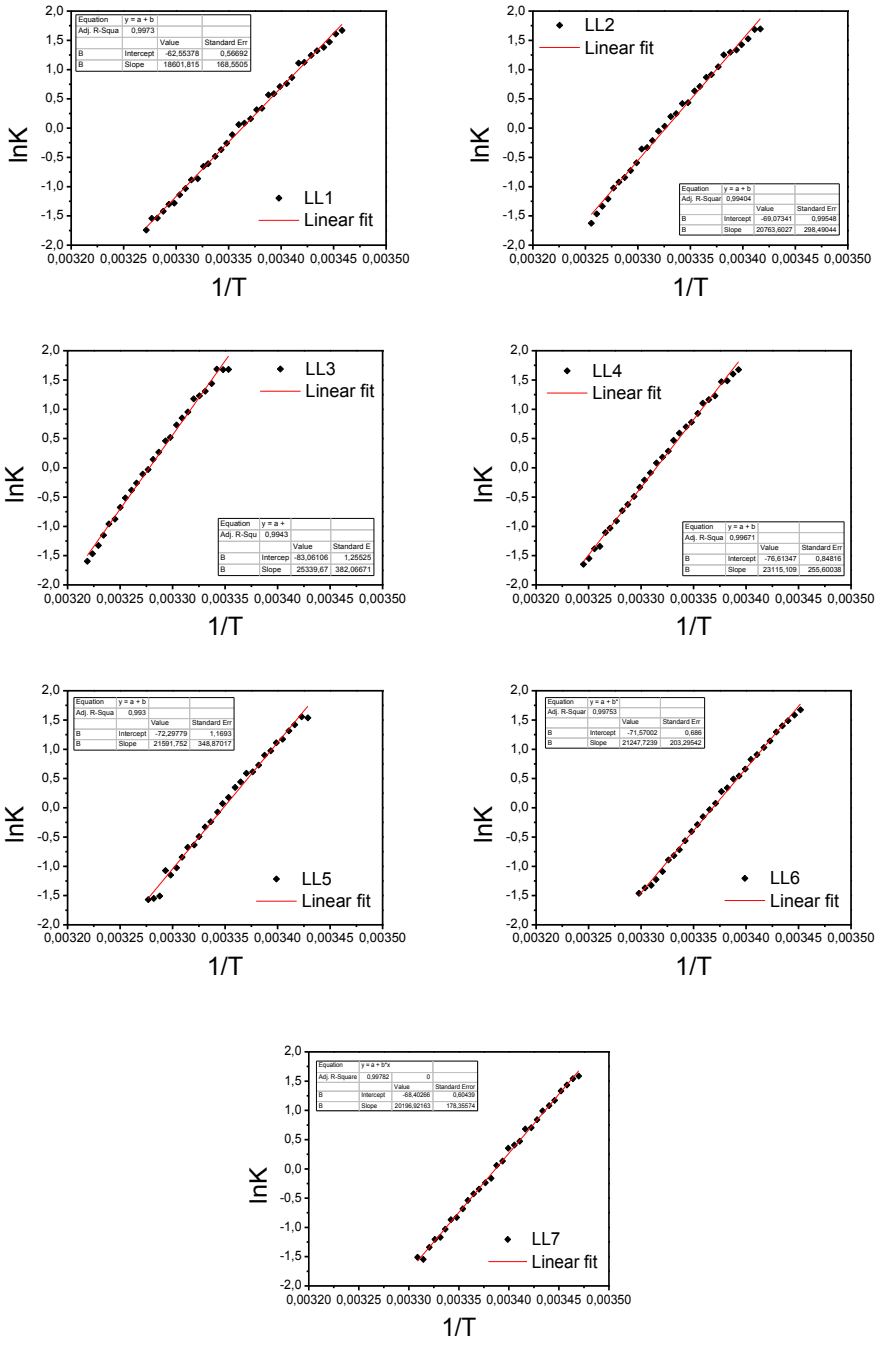


Figure 8-1. Van't Hoff plots for the sequences **LL1-7**.

8.8. NATIVE PAGE

20% polyacrylamide gels (acrylamide/bisacrylamide 19:1) of 0.75 mm were prepared in a Mini-PROTEAN Tetra Cell system from Bio-Rad. A 100 ml mother solution containing 38 g of acrylamide and 2 g of bisacrylamide was prepared (requires vigorous agitation), filtered (45 µm pore size) and stored at 4 ºC. The gels are then prepared *in situ* by adding the buffer solution and the polymerizing agents. For the preparation of the gels *in situ*, the following 10 ml solution was prepared for two gels: 5 ml acrylamide/bisacrylamide (19:1) mother solution, 2.5 ml NaPi buffer 0.2 M pH 7, 50 µl ammonium persulfate 20%, 10 µl TEMED and leveled up to 10 ml with water. This solution was added carefully between the glass plates with a syringe and the well comb was rapidly placed. The gels were then allowed to polymerize for half an hour. After removing the comb, the wells were rinsed gently with water. The glass plates containing the gel were fitted in the Mini-PROTEN device. A pre-running test was performed with dyes (bromophenol and cyanoxylenol solutions in glycerol/H_2O 1:3 v/v) in order to assess the good performance of the gels and the devices.

10 -20 µl samples containing 0.05 OD of the oligonucleotide were prepared in water/sucrose 2M 1:1 v/w and renatured as described in previous sections. After loading the samples and the poly-dT references (dT_{20}, dT_{25}, dT_{30}, dT_{35}, dT_{40} and dT_{50}) that were used as oligonucleotide length controls, the running of the gels was done for 6-7 h at 180 V in a thermally regulated chamber at 5 ºC. The visualization of the gels was achieved by treatment with Stains-All dye from Sigma-Aldrich (4 mg in formamide/H_2O 1:6 v/v) overnight. The Stains-All solution was afterwards rinsed with abundant water.

8.9. NMR

8.9.1.1. Sample preparation

The oligonucleotides, previously desalted, were dissolved in the adequate volume of the chosen solution in D_2O or H_2O/D_2O 9:1. The volumes were 0.5 ml for normal 5 mm tubes and 0.3 ml for narrow 3 mm tubes. The buffer solution used was 10 mM NaPi at different pHs with or wihout 100 mM NaCl. As internal reference DSS (sodium salt of the 4,4-dimethyl-4-silapentane-1-sulfonic acid perdeuterated) or TSP (sodium salt of the 3-(trimethylsilyl)-2,2',3,3'-tetradeuteropropionic acid) were used.

8.9.1.2. Spectra acquisition

All the oligonucleotide NMR spectra were recorded on Bruker AV-600 and Bruker AV-800 US2 devices from Spectroscopy and Molecular Structure Department of the “Instituto de Quimica-Fisica Rocasolano”, CSIC, Madrid.

The 2D spectra in D_2O were recorded mostly with 2k for dimension t_2 and 400 or 12 t_1 increments. In order to remove the water signal presaturation was used. The accumulations vary for every sample.

The NOESY experiments in D_2O were done with 100, 200 and 300 ms mixing time. The repetition time between accumulations was 1 – 3 s to allow the relaxation of the system.

The TOCSY experiments were done with 80 ms mixing time.

A module WATERGATE was used to remove the solvent signal during the pulse sequence before the acquisition. The mixing time varied from 50 to 300 ms.

8.9.1.3. Spectra assignment

Due to the high pH-sensitivity of the samples, all the NMR spectra have been acquired in H_2O/D_2O 90:10 conditions in order to avoid differences in the resulting pH values after adjustments of the different samples. In this conditions, signals corresponding to H3', H4' and H5'/H5" exhibit poor quality or are lost. Hence in some sequences these protons could not be completely assigned. For assigning exchangeable protons wide spectral windows up to 17 ppm are required. In this conditions non exchangeable protons region exhibits high overlapping and prevents specific assignment of most of the signals. For assigning non-exchangeable protons, hence, additional NOESY spectra with narrow spectral window were acquired.

After processing the spectra with the TopSpin software from Bruker, the assignment of the spectra were achieved as described in Chapter 0 using as a visualizing and assignment tool the software Sparky.[256] Cross-peaks of the NOESY spectra have been integrated and isolated cytosine H5-H6 and thymine H6-Me signals have been used to change intensities into distances. In cases of high overlapped signals, loose distances have been considered.

8.10. STRUCTURES CALCULATION

Initial calculations were performed with qualitative distance constraints (classified as 3, 4 or 5) and the resulting structures were then refined by employing more accurate distance constraints obtained from a complete relaxation matrix analysis with the program MARDIGRAS. Error bounds in the interprotonic distances were estimated by carrying out several MARDIGRAS calculations with different initial models, mixing times and correlation times. In addition to these experimentally derived constraints, hydrogen bond restraints were used. Target values for distances and angles related to hydrogen bonds were set to values obtained from crystallographic data in related structures.

Structures were calculated with the program DYANA 1.4 and further refined with the SANDER module of the molecular dynamics package AMBER 7.0. Initial DYANA calculations were carried out on the basis of qualitative distance constraints. The resulting structures were used as initial models in the complete relaxation matrix calculations to obtain accurate distance constraints, as described in the previous paragraph. These structures were taken as starting points for the AMBER refinement, consisting of an annealing protocol in vacuo, followed by long trajectories where explicit solvent molecules were included and using the Particle Mesh Ewald method to evaluate long-range electrostatic interactions. The specific protocols for these calculations have been described elsewhere. The AMBER-98 force field was used to describe the DNA, and the TIP3P model was used to simulate water molecules. Analysis of the representative structures as well as the MD trajectories was carried out with the programs Curves V5.1 and MOLMOL.

REFERENCES

(1) Watson, J. D.; Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171* (4356), 737–738. https://doi.org/10.1038/171737a0.

(2) Frankiln, R. E.; Gosling, R. G. Molecular Configuration in Sodium Thymonucleate. *Nature* **1953**, *171* (4356), 740–741. https://doi.org/10.1038/171740a0.

(3) Wilkins, M. H. F.; Stokes, A. R.; Wilson, H. R. Molecular Structure of Nucleic Acids: Molecular Structure of Deoxypentose Nucleic Acids. *Nature* **1953**, *171* (4356), 738–740. https://doi.org/10.1038/171738a0.

(4) Mirkin, S. M. Discovery of Alternative DNA Structures: A Heroic Decade (1979-1989). *Front. Biosci.* **2008**, *13*, 1064–1071.

(5) Zhao, J.; Bacolla, A.; Wang, G.; Vasquez, K. M. Non-B DNA Structure-Induced Genetic Instability and Evolution. *Cellular and Molecular Life Sciences*. 2010, pp 43–62. https://doi.org/10.1007/s00018-009-0131-2.

(6) Kaushik, M.; Kaushik, S.; Roy, K.; Singh, A.; Mahendru, S.; Kumar, M.; Chaudhary, S.; Ahmed, S.; Kukreti, S. A Bouquet of DNA Structures: Emerging Diversity. *Biochem. Biophys. Reports* **2016**, *5*, 388–395. https://doi.org/10.1016/j.bbrep.2016.01.013.

(7) Horton, N. C.; Finzel, B. C. The Structure of an RNA/DNA Hybrid: A Substrate of the Ribonuclease Activity of HIV-1 Reverse Transcriptase. *J. Mol. Biol.* **1996**, *264* (3), 521–533. https://doi.org/10.1006/jmbi.1996.0658.

(8) Jose, D.; Porschke, D. Dynamics of the B-A Transition of DNA Double Helices. *Nucleic Acids Res.* **2004**, *32* (7), 2251–2258. https://doi.org/10.1093/nar/gkh551.

(9) Wang, A. H.-J.; Quigley, G. J.; Kolpak, F. J.; Crawford, J. L.; Van Boom, J. H.; Van Der Marel, G.; Rich, A. Molecular Structure of a Left-Handed Double Helical DNA Fragment at Atomic Resolution. *Nature* **1979**, *282* (5740), 680–686. https://doi.org/10.1038/282680a0.

(10) Pohl, F. M.; Jovin, T. M. Salt-Induced Co-Operative Conformational Change of a Synthetic DNA: Equilibrium and Kinetic Studies with Poly(DG-DC). *J. Mol. Biol.* **1972**, *67* (3), 375–396. https://doi.org/10.1016/0022-2836(72)90457-3.

(11) Haniford, D. B.; Pulleyblank, D. E. The In-Vivo Occurrence of Z-DNA. *J. Biomol. Struct. Dyn.* **1983**, *1* (3), 593–609. https://doi.org/10.1080/07391102.1983.10507467.

(12) Wölfl, S.; Wittig, B.; Rich, A. Identification of Transcriptionally Induced Z-DNA Segments in the Human c-Myc Gene. *BBA - Gene Struct. Expr.* **1995**, *1264* (3), 294–302. https://doi.org/10.1016/0167-4781(95)00155-7.

(13) Herbert, A.; Lowenhaupt, K.; Spitzner, J.; Rich, A. Chicken Double-Stranded RNA Adenosine Deaminase Has Apparent Specificity for Z-DNA. *Proc. Natl. Acad. Sci.* **2006**, *92* (16), 7550–7554. https://doi.org/10.1073/pnas.92.16.7550.

(14) Rich, A.; Davies, D. R.; Crick, F. H. C.; Watson, J. D. The Molecular Structure of Polyadenylic Acid. *J. Mol. Biol.* **1961**, *3* (1), 71–86. https://doi.org/10.1016/S0022-2836(61)80009-0.

(15) Liu, K.; Miles, H. T.; Frazier, J.; Sasisekharan, V. A Novel DNA Duplex . A Parallel-Stranded DNA Helix with Hoogsteen Base Pairing A Novel DNA Duplex . A Parallel-Stranded DNA Helix with Hoogsteen Base Pairing. *Society* **2002**, *32* (May), 11802–11809. https://doi.org/10.1021/bi00095a008.

(16) Robinson, H.; Wang, A. H. 5'-CGA Sequence Is a Strong Motif for Homo Base-Paired Parallel-Stranded DNA Duplex as Revealed by NMR Analysis. *Proc. Natl. Acad. Sci.* **2006**, *90* (11), 5224–5228. https://doi.org/10.1073/pnas.90.11.5224.

(17) Szabat, M.; Kierzek, R. Parallel-Stranded DNA and RNA Duplexes – Structural Features and Potential Applications. *FEBS Journal*. John Wiley & Sons, Ltd (10.1111) December 1, 2017, pp 3986–3998. https://doi.org/10.1111/febs.14187.

(18) Hermann, T.; Patel, D. J. RNA Bulges as Architectural and Recognition Motifs. *Structure* **2000**, *8*, 47–54.

(19) Yoshizawa, S.; Kawai, G.; Watanabe, K.; Miura, K. I.; Hirao, I. *GNA Trinucleotide Loop Sequences Producing Extraordinarily Stable DNA Minihairpins*; 1997; Vol. 36. https://doi.org/10.1021/bi961738p.

(20) Svoboda, P.; Di Cara, A. Hairpin RNA: A Secondary Structure of Primary Importance. *Cellular and Molecular Life Sciences*. Birkhäuser-Verlag April 29, 2006, pp 901–918. https://doi.org/10.1007/s00018-005-5558-5.

(21) Felsenfeld, G.; Davies, D. R.; Rich, A. Formation of a Three-Stranded Polynucleotide Molecule. *Journal of the American Chemical Society*. American Chemical Society April 1957, pp 2023–2024. https://doi.org/10.1021/ja01565a074.

(22) De los Santos, C.; Rosen, M.; Patel, D. NMR Studies of DNA (R+)n.Cntdot.(Y-)n.Cntdot.(Y+)n Triple Helixes in Solution: Imino and Amino Proton Markers of T.Cntdot.A.Cntdot.T and C.Cntdot.G.Cntdot.C+ Base-Triple Formation. *Biochemistry* **2005**, *28* (18), 7282–7289. https://doi.org/10.1021/bi00444a021.

(23) Marck, C.; Thiele, D. Poly(DG).Poly(DC) at Neutral and Alkaline PH: The Formation of Triple Stranded Poly(DG).Poly(DG). Poly(DC). *Nucleic Acids Res.* **1978**, *5* (3), 1017–1028.

https://doi.org/10.1093/nar/5.3.1017.

(24) Kohwi, Y.; Kohwi-Shigematsu, T. Magnesium Ion-Dependent Triple-Helix Structure Formed by Homopurine-Homopyrimidine Sequences in Supercoiled Plasmid DNA. *Proc. Natl. Acad. Sci.* **1988**, *85* (11), 3781–3785. https://doi.org/10.1073/pnas.85.11.3781.

(25) Mirkin, S. M.; Lyamichev, V. I.; Drushlyak, K. N.; Dobrynin, V. N.; Filippov, S. A.; Frank-Kamenetskii, M. D. DNA H Form Requires a Homopurine-Homopyrimidine Mirror Repeat. *Nature* **1987**, *330* (6147), 495–497. https://doi.org/10.1038/330495a0.

(26) Jain, A.; Wang, G.; Vasquez, K. M. DNA Triple Helices: Biological Consequences and Therapeutic Potential. *Biochimie*. 2008, pp 1117–1130. https://doi.org/10.1016/j.biochi.2008.02.011.

(27) Gellert, M.; Lipsett, M. N.; Davies, D. R. Helix Formation by Guanylic Acid. *Proc. Natl. Acad. Sci. U. S. A.* **1962**, *48* (12), 2013–2018.

(28) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. Quadruplex DNA: Sequence, Topology and Structure. *Nucleic Acids Res.* **2006**, *34* (19), 5402–5415. https://doi.org/10.1093/nar/gkl655.

(29) Phan, A. T.; Patel, D. J. Two-Repeat Human Telomeric d(TAGGGTTAGGGT) Sequence Forms Interconverting Parallel and Antiparallel G-Quadruplexes in Solution: Distinct Topologies, Thermodynamic Properties, and Folding/Unfolding Kinetics. *J. Am. Chem. Soc.* **2003**, *125* (49), 15021–15027. https://doi.org/10.1021/ja037616j.

(30) Phan, A. T.; Modi, Y. S.; Patel, D. J. Two-Repeat Tetrahymena Telomeric d(TGGGGTTGGGGT) Sequence Interconverts between Asymmetric Dimeric G-Quadruplexes in Solution. *J. Mol. Biol.* **2004**, *338* (1), 93–102. https://doi.org/10.1016/j.jmb.2004.02.042.

(31) Schultze, P.; Smith, F. W.; Feigon, J. Refined Solution Structure of the Dimeric Quadruplex Formed from the Oxytricha Telomeric Oligonucleotide d(GGGGTTTTGGGG). *Structure* **1994**, *2* (3), 221–233. https://doi.org/10.1016/S0969-2126(00)00023-X.

(32) Rachwal, P. A.; Brown, T.; Fox, K. R. Effect of G-Tract Length on the Topology and Stability of Intramolecular DNA Quadruplexes. *Biochemistry* **2007**, *46* (11), 3036–3044. https://doi.org/10.1021/bi062118j.

(33) Risitano, A.; Fox, K. R. Influence of Loop Size on the Stability of Intramolecular DNA Quadruplexes. *Nucleic Acids Res.* **2004**, *32* (8), 2598–2606. https://doi.org/10.1093/nar/gkh598.

(34) Cevec, M.; Plavec, J. Role of Loop Residues and Cations on the Formation and Stability of

Dimeric DNA G-Quadruplexes. *Biochemistry* **2005**, *44* (46), 15238–15246. https://doi.org/10.1021/bi0514414.

(35) Dingley, A. J.; Peterson, R. D.; Grzesiek, S.; Feigon, J. Characterization of the Cation and Temperature Dependence of DNA Quadruplex Hydrogen Bond Properties Using High-Resolution NMR. *J. Am. Chem. Soc.* **2005**, *127* (41), 14466–14472. https://doi.org/10.1021/ja0540369.

(36) Kang, C.; Zhang, X.; Ratliff, R.; Moyzis, R.; Rich, A. Crystal Structure of Four-Stranded Oxytricha Telomeric DNA. *Nature* **1992**, *356* (6365), 126–131. https://doi.org/10.1038/356126a0.

(37) Sundquist, W. I.; Klug, A. Telomeric DNA Dimerizes by Formation of Guanine Tetrads between Hairpin Loops. *Nature* **1989**, *342* (6251), 825–829. https://doi.org/10.1038/342825a0.

(38) Sen, D.; Gilbert, W. Formation of Parallel Four-Stranded Complexes by Guanine-Rich Motifs in DNA and Its Implications for Meiosis. *Nature* **1988**, *334* (6180), 364–366. https://doi.org/10.1038/334364a0.

(39) Sun, D.; Hurley, L. H. The Importance of Negative Superhelicity in Inducing the Formation of G-Quadruplex and i-Motif Structures in the c-Myc Promoter: Implications for Drug Targeting and Control of Gene Expression. *J. Med. Chem.* **2009**, *52* (9), 2863–2874. https://doi.org/10.1021/jm900055s.

(40) Miyoshi, D.; Karimata, H.; Sugimoto, N. Hydration Regulates Thermodynamics of G-Quadruplex Formation under Molecular Crowding Conditions. *J. Am. Chem. Soc.* **2006**, *128* (24), 7957–7963. https://doi.org/10.1021/ja061267m.

(41) Lipps, H. J.; Rhodes, D. G-Quadruplex Structures: In Vivo Evidence and Function. *Trends in Cell Biology*. 2009, pp 414–422. https://doi.org/10.1016/j.tcb.2009.05.002.

(42) Huppert, J. L.; Balasubramanian, S. Prevalence of Quadruplexes in the Human Genome. *Nucleic Acids Res.* **2005**, *33* (9), 2908–2916. https://doi.org/10.1093/nar/gki609.

(43) Neidle, S.; Parkinson, G. Telomere Maintenance as a Target for Anticancer Drug Discovery. *Nature Reviews Drug Discovery*. European Association for Cardio-Thoracic Surgery 2002, pp 383–393. https://doi.org/10.1038/nrd793.

(44) Azzalin, C. M.; Reichenbach, P.; Khoriauli, L.; Giulotto, E.; Lingner, J. Telomeric Repeat-Containing RNA and RNA Surveillance Factors at Mammalian Chromosome Ends. *Science (80-.).* **2007**, *318* (5851), 798–801. https://doi.org/10.1126/science.1147182.

(45) Martadinata, H.; Phan, A. T. Structure of Human Telomeric RNA (TERRA): Stacking of Two

G-Quadruplex Blocks in K+ Solution. *Biochemistry* **2013**, *52* (13), 2176–2183. https://doi.org/10.1021/bi301606u.

(46) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative Visualization of DNA G-Quadruplex Structures in Human Cells. *Nat. Chem.* **2013**, *5* (3), 182–186. https://doi.org/10.1038/nchem.1548.

(47) Biffi, G.; Di Antonio, M.; Tannahill, D.; Balasubramanian, S. Visualization and Selective Chemical Targeting of RNA G-Quadruplex Structures in the Cytoplasm of Human Cells. *Nat. Chem.* **2014**, *6* (1), 75–80. https://doi.org/10.1038/nchem.1805.

(48) Brooks, T. A.; Kendrick, S.; Hurley, L. Making Sense of G-Quadruplex and i-Motif Functions in Oncogene Promoters. *FEBS Journal*. Blackwell Publishing Ltd September 1, 2010, pp 3459–3469. https://doi.org/10.1111/j.1742-4658.2010.07759.x.

(49) Bugaut, A.; Balasubramanian, S. 5′-UTR RNA G-Quadruplexes: Translation Regulation and Targeting. *Nucleic Acids Research*. 2012, pp 4727–4741. https://doi.org/10.1093/nar/gks068.

(50) Balasubramanian, S.; Hurley, L. H.; Neidle, S. Targeting G-Quadruplexes in Gene Promoters: A Novel Anticancer Strategy? *Nat. Rev. Drug Discov.* **2011**, *10* (4), 261–275. https://doi.org/10.1038/nrd3428.

(51) Double, J. A.; Burger, A. M.; Reszka, A. P.; Neidle, S.; Dai, F.; Moore, M. J.; Schultes, C. M. The G-Quadruplex-Interactive Molecule BRACO-19 Inhibits Tumor Growth, Consistent with Telomere Targeting and Interference with Telomerase Function. *Cancer Res.* **2005**, *65* (4), 1489–1496. https://doi.org/10.1158/0008-5472.can-04-2910.

(52) Mikami-Terao, Y.; Akiyama, M.; Yuza, Y.; Yanagisawa, T.; Yamada, O.; Kawano, T.; Agawa, M.; Ida, H.; Yamada, H. Antitumor Activity of TMPyP4 Interacting G-Quadruplex in Retinoblastoma Cell Lines. *Exp. Eye Res.* **2009**, *89* (2), 200–208. https://doi.org/10.1016/j.exer.2009.03.008.

(53) Müller, S.; Sanders, D. A.; Di Antonio, M.; Matsis, S.; Riou, J. F.; Rodriguez, R.; Balasubramanian, S. Pyridostatin Analogues Promote Telomere Dysfunction and Long-Term Growth Inhibition in Human Cancer Cells. *Org. Biomol. Chem.* **2012**, *10* (32), 6537–6546. https://doi.org/10.1039/c2ob25830g.

(54) Di Antonio, M.; Biffi, G.; Mariani, A.; Raiber, E. A.; Rodriguez, R.; Balasubramanian, S. Selective RNA versus DNA G-Quadruplex Targeting by Situ Click Chemistry. *Angew. Chemie - Int. Ed.* **2012**, *51* (44), 11073–11078. https://doi.org/10.1002/anie.201206281.

(55) Langridge, R.; Rich, A. Molecular Structure of Helical Polycytidylic Acid. *Nature* **1963**, *198*

(4882), 725–728. https://doi.org/10.1038/198725a0.

(56) Gehring, K.; Leroy, J.-L.; Guéron, M. A Tetrameric DNA Structure with Protonated Cytosine-Cytosine Base Pairs. *Nature* **1993**, *363* (6429), 561–565. https://doi.org/10.1038/363561a0.

(57) Dembska, A. The Analytical and Biomedical Potential of Cytosine-Rich Oligonucleotides: A Review. *Analytica Chimica Acta*. 2016, pp 1–12. https://doi.org/10.1016/j.aca.2016.05.007.

(58) Gallego, J.; Golden, E. B.; Stanley, D. E.; Reid, B. R. The Folding of Centromeric DNA Strands into Intercalated Structures: A Physicochemical and Computational Study. *J. Mol. Biol.* **1999**, *285* (3), 1039–1052. https://doi.org/10.1006/jmbi.1998.2334.

(59) Garavís, M.; Escaja, N.; Gabelica, V.; Villasante, A.; González, C. Centromeric Alpha-Satellite DNA Adopts Dimeric i-Motif Structures Capped by AT Hoogsteen Base Pairs. *Chem. - A Eur. J.* **2015**, *21* (27), 9816–9824. https://doi.org/10.1002/chem.201500448.

(60) Phan, A. T.; Guéron, M.; Leroy, J. L. The Solution Structure and Internal Motions of a Fragment of the Cytidine-Rich Strand of the Human Telomere. *J. Mol. Biol.* **2000**, *299* (1), 123–144. https://doi.org/10.1006/jmbi.2000.3613.

(61) Kendrick, S.; Akiyama, Y.; Hecht, S. M.; Hurley, L. H. The I-Motif in the Bcl-2 P1 Promoter Forms an Unexpectedly Stable Structure with a Unique 8:5:7 Loop Folding Pattern. *J. Am. Chem. Soc.* **2009**, *131* (48), 17667–17676. https://doi.org/10.1021/ja9076292.

(62) Dai, J.; Hatzakis, E.; Hurley, L. H.; Yang, D. I-Motif Structures Formed in the Human c-MYC Promoter Are Highly Dynamic-Insights into Sequence Redundancy and I-Motif Stability. *PLoS One* **2010**, *5* (7). https://doi.org/10.1371/journal.pone.0011647.

(63) Dzatko, S.; Krafcikova, M.; Hänsel-Hertsch, R.; Fessl, T.; Fiala, R.; Loja, T.; Krafcik, D.; Mergny, J. L.; Foldynova-Trantirkova, S.; Trantirek, L. Evaluation of the Stability of DNA I-Motifs in the Nuclei of Living Mammalian Cells. *Angew. Chemie - Int. Ed.* **2018**, *57* (8), 2165–2169. https://doi.org/10.1002/anie.201712284.

(64) Zeraati, M.; Langley, D. B.; Schofield, P.; Moye, A. L.; Rouet, R.; Hughes, W. E.; Bryan, T. M.; Dinger, M. E.; Christ, D. I-Motif DNA Structures Are Formed in the Nuclei of Human Cells. *Nat. Chem.* **2018**, *10* (6), 631–637. https://doi.org/10.1038/s41557-018-0046-3.

(65) Berger, I.; Egli, M.; Rich, A. Inter-Strand C-H...O Hydrogen Bonds Stabilizing Four-Stranded Intercalated Molecules: Stereoelectronic Effects of O4' in Cytosine-Rich DNA. *Proc. Natl. Acad. Sci.* **1996**, *93* (22), 12116–12121. https://doi.org/10.1073/pnas.93.22.12116.

(66) Phan, A. T.; Guéron, M.; Leroy, J.-L. The Solution Structure and Internal Motions of a Fragment of the Cytidine-Rich Strand of the Human Telomere. *J. Mol. Biol.* **2000**, *299* (1),

123–144. https://doi.org/10.1006/JMBI.2000.3613.

(67) Malliavin, T. E.; Gau, J.; Snoussi, K.; Leroy, J. L. Stability of the I-Motif Structure Is Related to the Interactions between Phosphodiester Backbones. *Biophys. J.* **2003**, *84* (6), 3838–3847. https://doi.org/10.1016/S0006-3495(03)75111-X.

(68) Mergny, J. L.; Lacroix, L.; Hélène, C.; Han, X.; Leroy, J. L. Intramolecular Folding of Pyrimidine Oligodeoxynucleotides into an I-DNA Motif. *J. Am. Chem. Soc.* **1995**, *117* (35), 8887–8898. https://doi.org/10.1021/ja00140a001.

(69) Zhou, J.; Wei, C.; Jia, G.; Wang, X.; Feng, Z.; Li, C. Formation of I-Motif Structure at Neutral and Slightly Alkaline PH. *Mol. BioSyst.* **2010**, *6* (3), 580–586. https://doi.org/10.1039/B919600E.

(70) Yang, B.; Rodgers, M. T. Base-Pairing Energies of Proton-Bound Heterodimers of Cytosine and Modified Cytosines: Implications for the Stability of DNA i-Motif Conformations. *J. Am. Chem. Soc.* **2014**, *136* (1), 282–290. https://doi.org/10.1021/ja409515v.

(71) Lieblein, A. L.; Krämer, M.; Dreuw, A.; Fürtig, B.; Schwalbe, H. The Nature of Hydrogen Bonds in Cytidine···H +··· Cytidine DNA Base Pairs. *Angew. Chemie - Int. Ed.* **2012**, *51* (17), 4067–4070. https://doi.org/10.1002/anie.201200549.

(72) Wright, E. P.; Huppert, J. L.; Waller, Z. A. E. Identification of Multiple Genomic DNA Sequences Which Form I-Motif Structures at Neutral PH. *Nucleic Acids Res.* **2017**, *45* (6), 2951–2959. https://doi.org/10.1093/nar/gkx090.

(73) Školáková, P.; Renčiuk, D.; Palacký, J.; Krafčík, D.; Dvořáková, Z.; Kejnovská, I.; Bednářová, K.; Vorlíčková, M. Systematic Investigation of Sequence Requirements for DNA I-Motif Formation. *Nucleic Acids Res.* **2019**, *47* (5), 2177–2189. https://doi.org/10.1093/nar/gkz046.

(74) Fleming, A. M.; Ding, Y.; Rogers, R. A.; Zhu, J.; Zhu, J.; Burton, A. D.; Carlisle, C. B.; Burrows, C. J. 4n-1 Is a “Sweet Spot” in DNA i-Motif Folding of 2′-Deoxycytidine Homopolymers. *J. Am. Chem. Soc.* **2017**, *139* (13), 4682–4689. https://doi.org/10.1021/jacs.6b10117.

(75) Brazier, J. A.; Shah, A.; Brown, G. D. I-Motif Formation in Gene Promoters: Unusually Stable Formation in Sequences Complementary to Known G-Quadruplexes. *Chem. Commun.* **2012**, *48* (87), 10739. https://doi.org/10.1039/c2cc30863k.

(76) Escaja, N.; Viladoms, J.; Garavís, M.; Villasante, A.; Pedroso, E.; González, C. A Minimal I-Motif Stabilized by Minor Groove G:T:G:T Tetrads. *Nucleic Acids Res.* **2012**, *40* (22), 11737–11747. https://doi.org/10.1093/nar/gks911.

(77) Brooks, T. A.; Kendrick, S.; Hurley, L. Making Sense of G-Quadruplex and i-Motif Functions in Oncogene Promoters. *FEBS Journal*. John Wiley & Sons, Ltd (10.1111) September 1, 2010, pp 3459–3469. https://doi.org/10.1111/j.1742-4658.2010.07759.x.

(78) Gurung, S. P.; Schwarz, C.; Hall, J. P.; Cardin, C. J.; Brazier, J. A. The Importance of Loop Length on the Stability of I-Motif Structures. *Chem. Commun.* **2015**, *51* (26), 5630–5632. https://doi.org/10.1039/c4cc07279k.

(79) Lieblein, A. L.; Fürtig, B.; Schwalbe, H. Optimizing the Kinetics and Thermodynamics of DNA I-Motif Folding. *ChemBioChem* **2013**, *14* (10), 1226–1230. https://doi.org/10.1002/cbic.201300284.

(80) Benabou, S.; Ferreira, R.; Aviñó, A.; González, C.; Lyonnais, S.; Solà, M.; Eritja, R.; Jaumot, J.; Gargallo, R. Solution Equilibria of Cytosine- and Guanine-Rich Sequences near the Promoter Region of the n-Myc Gene That Contain Stable Hairpins within Lateral Loops. *Biochim. Biophys. Acta - Gen. Subj.* **2014**, *1840* (1), 41–52. https://doi.org/10.1016/j.bbagen.2013.08.028.

(81) Benabou, S.; Garavís, M.; Lyonnais, S.; Eritja, R.; González, C.; Gargallo, R. Understanding the Effect of the Nature of the Nucleobase in the Loops on the Stability of the I-Motif Structure. *Phys. Chem. Chem. Phys.* **2016**, *18* (11), 7997–8004. https://doi.org/10.1039/c5cp07428b.

(82) McKim, M.; Buxton, A.; Johnson, C.; Metz, A.; Sheardy, R. D. Loop Sequence Context Influences the Formation and Stability of the I-Motif for DNA Oligomers of Sequence (CCCXXX)4, Where X = A and/or T, under Slightly Acidic Conditions. *J. Phys. Chem. B* **2016**, *120* (31), 7652–7661. https://doi.org/10.1021/acs.jpcb.6b04561.

(83) Canalia, M.; Leroy, J. L. [5mCCTCTCTCC]4: An i-Motif Tetramer with Intercalated T·T Pairs. *J. Am. Chem. Soc.* **2009**, *131* (36), 12870–12871. https://doi.org/10.1021/ja903210t.

(84) Kang, C.; Berger, I.; Lockshin, C.; Ratliff, R.; Moyzis, R.; Rich, A. Stable Loop in the Crystal Structure of the Intercalated Four-Stranded Cytosine-Rich Metazoan Telomere. *Proc. Natl. Acad. Sci.* **1995**, *92* (9), 3874–3878. https://doi.org/10.1073/pnas.92.9.3874.

(85) Esmaili, N.; Leroy, J. L. I-Motif Solution Structure and Dynamics of the d(AACCCC) and d(CCCCAA) Tetrahymena Telomeric Repeats. *Nucleic Acids Res.* **2005**, *33* (1), 213–224. https://doi.org/10.1093/nar/gki160.

(86) Gallego, J.; Chou, S. H.; Reid, B. R. Centromeric Pyrimidine Strands Fold into an Intercalated Motif by Forming a Double Hairpin with a Novel T:G:G:T Tetrad: Solution Structure of the d(TCCCGTTTCCA) Dimer. *J. Mol. Biol.* **1997**, *273* (4), 840–856. https://doi.org/10.1006/jmbi.1997.1361.

(87) Escaja, N.; Gelpí, J. L.; Orozco, M.; Rico, M.; Pedroso, E.; González, C. Four-Stranded DNA Structure Stabilized by a Novel G:C:A:T Tetrad. *J. Am. Chem. Soc.* **2003**, *125* (19), 5654–5662. https://doi.org/10.1021/ja0344157.

(88) Escaja, N.; Gómez-Pinto, I.; Pedroso, E.; González, C. Four-Stranded DNA Structures Can Be Stabilized by Two Different Types of Minor Groove G:C:G:C Tetrads. *J. Am. Chem. Soc.* **2007**, *129* (7), 2004–2014. https://doi.org/10.1021/ja066172z.

(89) Day, H. A.; Huguin, C.; Waller, Z. A. E. Silver Cations Fold I-Motif at Neutral PH. *Chem. Commun.* **2013**, *49* (70), 7696–7698. https://doi.org/10.1039/c3cc43495h.

(90) Abdelhamid, M. A.; Fábián, L.; Macdonald, C. J.; Cheesman, M. R.; Gates, A. J.; Waller, Z. A. Redox-Dependent Control of i-Motif DNA Structure Using Copper Cations. *Nucleic Acids Res.* **2018**, *46* (12), 5886–5893. https://doi.org/10.1093/nar/gky390.

(91) Miyoshi, D.; Matsumura, S.; Nakano, S. I.; Sugimoto, N. Duplex Dissociation of Telomere DNAs Induced by Molecular Crowding. *J. Am. Chem. Soc.* **2004**, *126* (1), 165–169. https://doi.org/10.1021/ja036721q.

(92) Cui, J.; Waltman, P.; Le, V. H.; Lewis, E. A. The Effect of Molecular Crowding on the Stability of Human C-MYC Promoter Sequence I-Motif at Neutral PH. *Molecules* **2013**, *18* (10), 12751–12767. https://doi.org/10.3390/molecules181012751.

(93) Saxena, S.; Joshi, S.; Shankaraswamy, J.; Tyagi, S.; Kukreti, S. Magnesium and Molecular Crowding of the Cosolutes Stabilize the I-Motif Structure at Physiological PH. *Biopolymers* **2017**, *107* (7), e23018. https://doi.org/10.1002/bip.23018.

(94) Zhao, C.; Ren, J.; Qu, X. Single-Walled Carbon Nanotubes Binding to Human Telomeric i-Motif DNA under Molecular-Crowding Conditions: More Water Molecules Released. *Chem. - A Eur. J.* **2008**, *14* (18), 5435–5439. https://doi.org/10.1002/chem.200800280.

(95) Li, X.; Peng, Y.; Ren, J.; Qu, X. Carboxyl-Modified Single-Walled Carbon Nanotubes Selectively Induce Human Telomeric i-Motif Formation. *Proc. Natl. Acad. Sci.* **2006**, *103* (52), 19658–19663. https://doi.org/10.1073/pnas.0607245103.

(96) Dvořáková, Z.; Renčiuk, D.; Kejnovská, I.; Školáková, P.; Bednářová, K.; Sagi, J.; Vorlíčková, M. I-Motif of Cytosine-Rich Human Telomere DNA Fragments Containing Natural Base Lesions. *Nucleic Acids Res.* **2018**, *46* (4), 1624–1634. https://doi.org/10.1093/nar/gky035.

(97) Xu, B.; Devi, G.; Shao, F. Regulation of Telomeric I-Motif Stability by 5-Methylcytosine and 5-Hydroxymethylcytosine Modification. *Org. Biomol. Chem.* **2015**, *13* (20), 5646–5651. https://doi.org/10.1039/C4OB02646B.

(98) Yang, B.; Rodgers, M. T. Base-Pairing Energies of Protonated Nucleoside Base Pairs of DCyd and M5dCyd: Implications for the Stability of DNA i-Motif Conformations. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (8), 1394–1403. https://doi.org/10.1007/s13361-015-1144-8.

(99) Yang, B.; Wu, R. R.; Rodgers, M. T. Base-Pairing Energies of Proton-Bound Homodimers Determined by Guided Ion Beam Tandem Mass Spectrometry: Application to Cytosine and 5-Substituted Cytosines. *Anal. Chem.* **2013**, *85* (22), 11000–11006. https://doi.org/10.1021/ac402542g.

(100) Tsvetkov, V. B.; Zatsepin, T. S.; Belyaev, E. S.; Kostyukevich, Y. I.; Shpakovski, G. V; Podgorsky, V. V; Pozmogova, G. E.; Varizhuk, A. M.; Aralov, A. V. I-Clamp Phenoxazine for the Fine Tuning of DNA i-Motif Stability. *Nucleic Acids Res.* **2018**, *46* (6), 2751–2764. https://doi.org/10.1093/nar/gky121.

(101) Collin, D.; Gehring, K. Stability of Chimeric DNA/RNA Cytosine Tetrads: Implications for i-Motif Formation by RNA. *J. Am. Chem. Soc.* **1998**, *120* (17), 4069–4072. https://doi.org/10.1021/ja973346r.

(102) Snoussi, K.; Nonin-Lecomte, S.; Leroy, J. L. The RNA I-Motif. *J. Mol. Biol.* **2001**, *309* (1), 139–153. https://doi.org/10.1006/jmbi.2001.4618.

(103) Kumar, N.; Nielsen, J. T.; Maiti, S.; Petersen, M. I-Motif Formation with Locked Nucleic Acid (LNA). *Angew. Chemie - Int. Ed.* **2007**, *46* (48), 9220–9222. https://doi.org/10.1002/anie.200701667.

(104) Fenna, C. P.; Wilkinson, V. J.; Arnold, J. R. P.; Cosstick, R.; Fisher, J. The Effect of 2′-Fluorine Substitutions on DNA i-Motif Conformation and Stability. *Chem. Commun.* **2008**, No. 30, 3567–3569. https://doi.org/10.1039/b804833a.

(105) Robidoux, S.; Damha, M. J. D-2-Deoxyribose and d-Arabinose, but Not d-Ribose, Stabilize the Cytosine Tetrad (i-Dna) Structure. *J. Biomol. Struct. Dyn.* **1997**, *15* (3), 529–535. https://doi.org/10.1080/07391102.1997.10508963.

(106) Assi, H. A.; Harkness, R. W.; Martin-Pintado, N.; Wilds, C. J.; Campos-Olivas, R.; Mittermaier, A. K.; González, C.; Damha, M. J. Stabilization of I-Motif Structures by 2'-β-Fluorination of DNA. *Nucleic Acids Res.* **2016**, *44* (11), 4998–5009. https://doi.org/10.1093/nar/gkw402.

(107) Aviñó, A.; Dellafiore, M.; Gargallo, R.; González, C.; Iribarren, A. M.; Montserrat, J.; Eritja, R. Stabilization of Telomeric I-Motif Structures by (2′S)-2′-Deoxy-2′-C-Methylcytidine Residues. *ChemBioChem* **2017**, *18* (12), 1123–1128. https://doi.org/10.1002/cbic.201700112.

(108) Abou Assi, H.; Lin, Y. C.; Serrano, I.; González, C.; Damha, M. J. Probing Synergistic Effects

of DNA Methylation and 2′-β-Fluorination on i-Motif Stability. *Chem. - A Eur. J.* **2018**, *24* (2), 471–477. https://doi.org/10.1002/chem.201704591.

(109) Mergny, J. L.; Lacroix, L. Kinetics and Thermodynamics of I-DNA Formation: Phosphodiester versus Modified Oligodeoxynucleotides. *Nucleic Acids Res.* **1998**, *26* (21), 4797–4803. https://doi.org/10.1093/nar/26.21.4797.

(110) Krishnan-Ghosh, Y.; Stephens, E.; Balasubramanian, S. PNA Forms an I-Motif. *Chem. Commun.* **2005**, No. 42, 5278–5280. https://doi.org/10.1039/b510405j.

(111) Modi, S.; Wani, A. H.; Krishnan, Y. The PNA-DNA Hybrid I-Motif: Implications for Sugar-Sugar Contacts in i-Motif Tetramerization. *Nucleic Acids Res.* **2006**, *34* (16), 4354–4363. https://doi.org/10.1093/nar/gkl443.

(112) Dong, Y.; Yang, Z.; Liu, D. DNA Nanotechnology Based on I-Motif Structures. *Acc. Chem. Res.* **2014**, *47* (6), 1853–1860. https://doi.org/10.1021/ar500073a.

(113) Shi, L.; Peng, P.; Du, Y.; Li, T. Programmable I-Motif DNA Folding Topology for a PH-Switched Reversible Molecular Sensing Device. *Nucleic Acids Res.* **2017**, *45* (8), 4306–4314. https://doi.org/10.1093/nar/gkx202.

(114) Protopopova, A. D.; Tsvetkov, V. B.; Varizhuk, A. M.; Barinov, N. A.; Podgorsky, V. V; Klinov, D. V; Pozmogova, G. E. The Structural Diversity of C-Rich DNA Aggregates: Unusual Self-Assembly of Beetle-like Nanostructures. *Phys. Chem. Chem. Phys.* **2018**, *20* (5), 3543–3553. https://doi.org/10.1039/c7cp05380k.

(115) Lu, C.; Huang, Z.; Liu, B.; Liu, Y.; Ying, Y.; Liu, J. Poly-Cytosine DNA as a High-Affinity Ligand for Inorganic Nanomaterials. *Angew. Chemie - Int. Ed.* **2017**, *56* (22), 6208–6212. https://doi.org/10.1002/anie.201702998.

(116) Tanaka, S.; Yukami, S.; Fukushima, K.; Wakabayashi, K.; Ohya, Y.; Kuzuya, A. Bulk PH-Responsive DNA Quadruplex Hydrogels Prepared by Liquid-Phase, Large-Scale DNA Synthesis. *ACS Macro Lett.* **2018**, *7* (3), 295–299. https://doi.org/10.1021/acsmacrolett.8b00063.

(117) Deshpande, S. R.; Hammink, R.; Nelissen, F. H. T.; Rowan, A. E.; Heus, H. A. Biomimetic Stress Sensitive Hydrogel Controlled by DNA Nanoswitches. *Biomacromolecules* **2017**, *18* (10), 3310–3317. https://doi.org/10.1021/acs.biomac.7b00964.

(118) Heinen, L.; Walther, A. Temporal Control of I-Motif Switch Lifetimes for Autonomous Operation of Transient DNA Nanostructures. *Chem. Sci.* **2017**, *8* (5), 4100–4107. https://doi.org/10.1039/c7sc00646b.

(119) Zhu, J.; Kim, Y.; Lin, H.; Wang, S.; Mirkin, C. A. PH-Responsive Nanoparticle Superlattices with Tunable DNA Bonds. *J. Am. Chem. Soc.* **2018**, *140* (15), 5061–5064. https://doi.org/10.1021/jacs.8b02793.

(120) Huang, J.; Ying, L.; Yang, X.; Yang, Y.; Quan, K.; Wang, H.; Xie, N.; Ou, M.; Zhou, Q.; Wang, K. Ratiometric Fluorescent Sensing of PH Values in Living Cells by Dual-Fluorophore-Labeled i-Motif Nanoprobes. *Anal. Chem.* **2015**, *87* (17), 8724–8731. https://doi.org/10.1021/acs.analchem.5b01527.

(121) Ma, W.; Yan, L.; He, X.; Qing, T.; Lei, Y.; Qiao, Z.; He, D.; Huang, K.; Wang, K. Hairpin-Contained i-Motif Based Fluorescent Ratiometric Probe for High-Resolution and Sensitive Response of Small PH Variations. *Anal. Chem.* **2018**, *90* (3), 1889–1896. https://doi.org/10.1021/acs.analchem.7b03972.

(122) Ying, L.; Xie, N.; Yang, Y.; Yang, X.; Zhou, Q.; Yin, B.; Huang, J.; Wang, K. A Cell-Surface-Anchored Ratiometric i-Motif Sensor for Extracellular PH Detection. *Chem. Commun.* **2016**, *52* (50), 7818–7821. https://doi.org/10.1039/c6cc03163c.

(123) Xu, B.; Wu, X.; Yeow, E. K. L.; Shao, F. A Single Thiazole Orange Molecule Forms an Exciplex in a DNA I-Motif. *Chem. Commun.* **2014**, *50* (48), 6402–6405. https://doi.org/10.1039/c4cc01147c.

(124) Li, L.; Jiang, Y.; Cui, C.; Yang, Y.; Zhang, P.; Stewart, K.; Pan, X.; Li, X.; Yang, L.; Qiu, L.; et al. Modulating Aptamer Specificity with PH-Responsive DNA Bonds. *J. Am. Chem. Soc.* **2018**, *140* (41), 13335–13339. https://doi.org/10.1021/jacs.8b08047.

(125) Lu, J.; Sun, J.; Li, F.; Wang, J.; Liu, J.; Kim, D.; Fan, C.; Hyeon, T.; Ling, D. Highly Sensitive Diagnosis of Small Hepatocellular Carcinoma Using PH-Responsive Iron Oxide Nanocluster Assemblies. *J. Am. Chem. Soc.* **2018**, *140* (32), 10071–10074. https://doi.org/10.1021/jacs.8b04169.

(126) Huang, F.; Liao, W. C.; Sohn, Y. S.; Nechushtai, R.; Lu, C. H.; Willner, I. Light-Responsive and PH-Responsive DNA Microcapsules for Controlled Release of Loads. *J. Am. Chem. Soc.* **2016**, *138* (28), 8936–8945. https://doi.org/10.1021/jacs.6b04773.

(127) Park, H.; Kim, J.; Jung, S.; Kim, W. J. DNA-Au Nanomachine Equipped with i-Motif and G-Quadruplex for Triple Combinatorial Anti-Tumor Therapy. *Adv. Funct. Mater.* **2018**, *28* (5). https://doi.org/10.1002/adfm.201705416.

(128) Jang, D.; Lee, Y. M.; Lee, J.; Doh, J.; Kim, W. J. Remission of Lymphoblastic Leukaemia in an Intravascular Fluidic Environment by Pliable Drug Carrier with a Sliding Target Ligand. *Sci. Rep.* **2017**, *7*. https://doi.org/10.1038/srep40739.

(129) Chen, Y.; Qu, K.; Zhao, C.; Wu, L.; Ren, J.; Wang, J.; Qu, X. Insights into the Biomedical Effects of Carboxylated Single-Wall Carbon Nanotubes on Telomerase and Telomeres. *Nat. Commun.* **2012**, *3*. https://doi.org/10.1038/ncomms2091.

(130) Kasinathan, S.; Henikoff, S. Non-B-Form DNA Is Enriched at Centromeres. *Mol. Biol. Evol.* **2018**, *35* (4), 949–962. https://doi.org/10.1093/molbev/msy010.

(131) Alexandrov, I.; Kazakov, A.; Tumeneva, I.; Shepelev, V.; Yurov, Y. Alpha-Satellite DNA of Primates: Old and New Families. *Chromosoma* **2001**, *110* (4), 253–266.

(132) Garavís, M.; Escaja, N.; Gabelica, V.; Villasante, A.; González, C. Centromeric Alpha-Satellite DNA Adopts Dimeric i-Motif Structures Capped by at Hoogsteen Base Pairs. *Chem. - A Eur. J.* **2015**, *21* (27), 9816–9824. https://doi.org/10.1002/chem.201500448.

(133) Garavís, M.; Méndez-Lago, M.; Gabelica, V.; Whitehead, S. L.; González, C.; Villasante, A. The Structure of an Endogenous Drosophila Centromere Reveals the Prevalence of Tandemly Repeated Sequences Able to Form I-Motifs. *Sci. Rep.* **2015**, *5*. https://doi.org/10.1038/srep13307.

(134) Kendrick, S.; Kang, H. J.; Alam, M. P.; Madathil, M. M.; Agrawal, P.; Gokhale, V.; Yang, D.; Hecht, S. M.; Hurley, L. H. The Dynamic Character of the BCL2 Promoter I-Motif Provides a Mechanism for Modulation of Gene Expression by Compounds That Bind Selectively to the Alternative DNA Hairpin Structure. *J. Am. Chem. Soc.* **2014**, *136* (11), 4161–4171. https://doi.org/10.1021/ja410934b.

(135) Kang, H.-J.; Kendrick, S.; Hecht, S. M.; Hurley, L. H. The Transcriptional Complex Between the BCL2 I-Motif and HnRNP LL Is a Molecular Switch for Control of Gene Expression That Can Be Modulated by Small Molecules. **2014**. https://doi.org/10.1021/ja4109352.

(136) Sutherland, C.; Cui, Y.; Mao, H.; Hurley, L. H. A Mechanosensor Mechanism Controls the G-Quadruplex/i-Motif Molecular Switch in the MYC Promoter NHE III1. *J. Am. Chem. Soc.* **2016**, *138* (42), 14138–14151. https://doi.org/10.1021/jacs.6b09196.

(137) Brown, R.; Wang, T.; Chappeta, V. R.; Wu, G.; Onel, B.; Chawla, R.; Quijada, H.; Camp, S. M.; Chiang, E. T.; Lassiter, Q.; et al. The Consequences of Overlapping G-Quadruplexes and i-Motifs in the PDGFR- # Core Promoter NHE Can Explain the Unexpected Effects of Mutations and Provide Opportunities for Selective Targeting of Both Structures by Small Molecules to Downregulate Gene Expr. **2017**. https://doi.org/10.1021/jacs.6b10028.

(138) Castellano, E.; Santos, E. Functional Specificity of Ras Isoforms: So Similar but so Different. *Genes and Cancer* **2011**, *2* (3), 216–231. https://doi.org/10.1177/1947601911408081.

(139) Kaiser, C. E.; Van Ert, N. A.; Agrawal, P.; Chawla, R.; Yang, D.; Hurley, L. H. Insight into the Complexity of the I-Motif and G-Quadruplex DNA Structures Formed in the KRAS Promoter and Subsequent Drug-Induced Gene Repression. *J. Am. Chem. Soc.* **2017**, *139* (25), 8522–8536. https://doi.org/10.1021/jacs.7b02046.

(140) Miglietta, G.; Cogoi, S.; Pedersen, E. B.; Xodo, L. E. GC-Elements Controlling HRAS Transcription Form i-Motif Structures Unfolded by Heterogeneous Ribonucleoprotein Particle A1. *Sci. Rep.* **2015**, *5*, 18097. https://doi.org/10.1038/srep18097.

(141) Uribe, D. J.; Guo, K.; Shin, Y. J.; Sun, D. Heterogeneous Nuclear Ribonucleoprotein K and Nucleolin as Transcriptional Activators of the Vascular Endothelial Growth Factor Promoter through Interaction with Secondary DNA Structures. *Biochemistry* **2011**, *50* (18), 3796–3806. https://doi.org/10.1021/bi101633b.

(142) Day, H. A.; Pavlou, P.; Waller, Z. A. E. I-Motif DNA: Structure, Stability and Targeting with Ligands. *Bioorg. Med. Chem.* **2014**, *22* (16), 4407–4418. https://doi.org/10.1016/J.BMC.2014.05.047.

(143) Debnath, M.; Ghosh, S.; Chauhan, A.; Paul, R.; Bhattacharyya, K.; Dash, J. Preferential Targeting of I-Motifs and G-Quadruplexes by Small Molecules. *Chem. Sci.* **2017**, *8* (11), 7448–7456. https://doi.org/10.1039/c7sc02693e.

(144) Sedghi Masoud, S.; Yamaoki, Y.; Ma, Y.; Marchand, A.; Winnerdy, F. R.; Gabelica, V.; Phan, A. T.; Katahira, M.; Nagasawa, K. Analysis of Interactions between Telomeric I-Motif DNA and a Cyclic Tetraoxazole Compound. *ChemBioChem* **2018**, *19* (21), 2268–2272. https://doi.org/10.1002/cbic.201800425.

(145) Shu, B.; Cao, J.; Kuang, G.; Qiu, J.; Zhang, M.; Zhang, Y.; Wang, M.; Li, X.; Kang, S.; Ou, T. M.; et al. Syntheses and Evaluation of New Acridone Derivatives for Selective Binding of Oncogene C-: Myc Promoter i-Motifs in Gene Transcriptional Regulation. *Chem. Commun.* **2018**, *54* (16), 2036–2039. https://doi.org/10.1039/c8cc00328a.

(146) Wright, E. P.; Day, H. A.; Ibrahim, A. M.; Kumar, J.; Boswell, L. J. E.; Huguin, C.; Stevenson, C. E. M.; Pors, K.; Waller, Z. A. E. Mitoxantrone and Analogues Bind and Stabilize I-Motif Forming DNA Sequences. *Sci. Rep.* **2016**, *6* (1), 39456. https://doi.org/10.1038/srep39456.

(147) Journey, S. N.; Alden, S. L.; Hewitt, W. M.; Peach, M. L.; Nicklaus, M. C.; Schneekloth, J. S. Probing the: Hras -1 Y i-Motif with Small Molecules. *Medchemcomm* **2018**, *9* (12), 2000–2007. https://doi.org/10.1039/c8md00311d.

(148) Fedoroff, O. Y.; Rangan, A.; Chemeris, V. V; Hurley, L. H. Cationic Porphyrins Promote the Formation of I-Motif DNA and Bind Peripherally by a Nonintercalative Mechanism.

Biochemistry **2000**, *39* (49), 15083–15090. https://doi.org/10.1021/bi001528j.

(149) Lee, I. J.; P. Patil, S.; Fhayli, K.; Alsaiari, S.; Khashab, N. M. Probing Structural Changes of Self Assembled I-Motif DNA. *Chem. Commun.* **2015**, *51* (18), 3747–3749. https://doi.org/10.1039/c4cc06824f.

(150) Pagano, A.; Iaccarino, N.; Abdelhamid, M. A. S.; Brancaccio, D.; Garzarella, E. U.; Di Porzio, A.; Novellino, E.; Waller, Z. A. E.; Pagano, B.; Amato, J.; et al. Common G-Quadruplex Binding Agents Found to Interact With i-Motif-Forming DNA: Unexpected Multi-Target-Directed Compounds. *Front. Chem.* **2018**, *6*, 281. https://doi.org/10.3389/fchem.2018.00281.

(151) Abdelhamid, M. A. S.; Gates, A. J.; Waller, Z. A. E. Destabilization of I-Motif DNA at Neutral PH by G-Quadruplex Ligands. *Biochemistry* **2019**, *58* (4), 245–249. https://doi.org/10.1021/acs.biochem.8b00968.

(152) Pullman, B.; Jortner, J. IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN); Springer, Dordrecht, 1983; pp 559–565. https://doi.org/10.1007/978-94-009-7225-4_43.

(153) Altona, C.; Sundaralingam, M. Conformational Analysis of the Sugar Ring in Nucleosides and Nucleotides. a New Description Using the Concept of Pseudorotation. *J. Am. Chem. Soc.* **1972**, *94* (23), 8205–8212. https://doi.org/10.1021/ja00778a043.

(154) DeVoe, H.; Tinoco, I. The Stability of Helical Polynucleotides: Base Contributions. *J. Mol. Biol.* **1962**, *4* (6), 500–517. https://doi.org/10.1016/S0022-2836(62)80105-3.

(155) Yakovchuk, P.; Protozanova, E.; Frank-Kamenetskii, M. D. Base-Stacking and Base-Pairing Contributions into Thermal Stability of the DNA Double Helix. *Nucleic Acids Res.* **2006**, *34* (2), 564. https://doi.org/10.1093/NAR/GKJ454.

(156) Kollman, P. A.; Allen, L. C. Theory of the Hydrogen Bond. *Chem. Rev.* **1972**, *72* (3), 283–303. https://doi.org/10.1021/cr60277a004.

(157) Lemieux, S.; Major, F. RNA Canonical and Non-Canonical Base Pairing Types: A Recognition Method and Complete Repertoire. *Nucleic Acids Res.* **2002**, *30* (19), 4250–4263. https://doi.org/10.1093/nar/gkf540.

(158) Nikolova, E. N.; Kim, E.; Wise, A. A.; O'Brien, P. J.; Andricioaei, I.; Al-Hashimi, H. M. Transient Hoogsteen Base Pairs in Canonical Duplex DNA. *Nature* **2011**, *470* (7335), 498–504. https://doi.org/10.1038/nature09775.

(159) Chapter 11. Physical and Structural Techniques Applied to Nucleic Acids. In *Nucleic Acids in Chemistry and Biology*; Royal Society of Chemistry: Cambridge, 2007; pp 427–458. https://doi.org/10.1039/9781847555380-00427.

(160) DeVoe, H.; Tinoco, I. The Hypochromism of Helical Polynucleotides. *J. Mol. Biol.* **1962**, *4* (6), 518–527. https://doi.org/10.1016/S0022-2836(62)80106-5.

(161) Gray, D. M.; Ratliff, R. L.; Vaughan, M. R. Circular Dichroism Spectroscopy of DNA. *Methods Enzymol.* **1992**, *211* (C), 389–406. https://doi.org/10.1016/0076-6879(92)11021-A.

(162) Bishop, G. R.; Chaires, J. B. Characterization of DNA Structures by Circular Dichroism. In *Current Protocols in Nucleic Acid Chemistry*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2002; Vol. 11, pp 7.11.1-7.11.8. https://doi.org/10.1002/0471142700.nc0711s11.

(163) Kypr, J.; Kejnovská, I.; Renčiuk, D.; Vorlíčková, M. Circular Dichroism and Conformational Polymorphism of DNA. *Nucleic Acids Research*. Oxford University Press April 2009, pp 1713–1725. https://doi.org/10.1093/nar/gkp026.

(164) Xu, W.; Chan, K. M.; Kool, E. T. Fluorescent Nucleobases as Tools for Studying DNA and RNA. *Nature Chemistry*. Nature Publishing Group November 16, 2017, pp 1043–1055. https://doi.org/10.1038/NCHEM.2859.

(165) Didenko, V. V. DNA Probes Using Fluoresence Resonance Energy Transfer. *Biotechniques* **2001**, *31* (5), 1106–1121. https://doi.org/10.2144/01315rv02.

(166) Wang, K.; Tang, Z.; Yang, C. J.; Kim, Y.; Fang, X.; Li, W.; Wu, Y.; Medley, C. D.; Cao, Z.; Li, J.; et al. Molecular Engineering of DNA: Molecular Beacons. *Angewandte Chemie - International Edition*. John Wiley & Sons, Ltd January 19, 2009, pp 856–870. https://doi.org/10.1002/anie.200800370.

(167) Sheng, Q.; Neaverson, J. C.; Mahmoud, T.; Stevenson, C. E. M.; Matthews, S. E.; Waller, Z. A. E. Identification of New DNA I-Motif Binding Ligands through a Fluorescent Intercalator Displacement Assay. *Org. Biomol. Chem.* **2017**, *15* (27), 5669–5673. https://doi.org/10.1039/c7ob00710h.

(168) Houseal, T. W.; Bustamante, C.; Stump, R. F.; Maestre, M. F. Real-Time Imaging of Single DNA Molecules with Fluorescence Microscopy. *Biophys. J.* **1989**, *56* (3), 507–516. https://doi.org/10.1016/S0006-3495(89)82697-9.

(169) Suzuki, T.; Fujikura, K.; Higashiyama, T.; Takata, K. DNA Staining for Fluorescence and Laser Confocal Microscopy. *J. Histochem. Cytochem.* **1997**, *45* (1), 49–53. https://doi.org/10.1177/002215549704500107.

(170) Zeraati, M.; Langley, D. B.; Schofield, P.; Moye, A. L.; Rouet, R.; Hughes, W. E.; Bryan, T. M.; Dinger, M. E.; Christ, D. I-Motif DNA Structures Are Formed in the Nuclei of Human Cells. *Nat. Chem.* **2018**, *10* (6), 631–637. https://doi.org/10.1038/s41557-018-0046-3.

(171) Mergny, J. L.; Lacroix, L. Analysis of Thermal Melting Curves. *Oligonucleotides*. Mary Ann Liebert, Inc. December 8, 2003, pp 515–537. https://doi.org/10.1089/154545703322860825.

(172) Wüthrich, K. NMR of Proteins and Nucleic Acids. 1986.

(173) James, T. L. NMR Determination of Oligonucleotide Structure. *Curr. Protoc. Nucleic Acid Chem.* **2000**, *00* (1), 7.2.1-7.2.16. https://doi.org/10.1002/0471142700.nc0702s00.

(174) Braunschweiler, L.; Ernst, R. R. Coherence Transfer by Isotropic Mixing: Application to Proton Correlation Spectroscopy. *J. Magn. Reson.* **1983**, *53* (3), 521–528. https://doi.org/10.1016/0022-2364(83)90226-3.

(175) Kumar, A.; Ernst, R. R.; Wüthrich, K. A Two-Dimensional Nuclear Overhauser Enhancement (2D NOE) Experiment for the Elucidation of Complete Proton-Proton Cross-Relaxation Networks in Biological Macromolecules. *Biochem. Biophys. Res. Commun.* **1980**, *95* (1), 1–6. https://doi.org/10.1016/0006-291X(80)90695-6.

(176) Scheek, R. M.; Russo, N.; Boelens, R.; Kaptein, R.; van Boom, J. H. Sequential Resonance Assignments in DNA 1H NMR Spectra by Two-Dimensional NOE Spectroscopy. *J. Am. Chem. Soc.* **1983**, *105* (9), 2914–2916. https://doi.org/10.1021/ja00347a075.

(177) Hare, D. R.; Wemmer, D. E.; Chou, S. H.; Drobny, G.; Reid, B. R. Assignment of the Non-Exchangeable Proton Resonances of d(C-G-C-G-A-A-T-T-C-G-C-G) Using Two-Dimensional Nuclear Magnetic Resonance Methods. *J. Mol. Biol.* **1983**, *171* (3), 319–336. https://doi.org/10.1016/0022-2836(83)90096-7.

(178) Keepers, J. W.; James, T. L. A Theoretical Study of Distance Determinations from NMR. Two-Dimensional Nuclear Overhauser Effect Spectra. *J. Magn. Reson.* **1984**, *57* (3), 404–426. https://doi.org/10.1016/0022-2364(84)90257-9.

(179) Borgias, B. A.; James, T. L. Two-Dimensional Nuclear Overhauser Effect: Complete Relaxation Matrix Analysis. *Methods Enzymol.* **1989**, *176* (C), 169–183. https://doi.org/10.1016/0076-6879(89)76011-0.

(180) Borgias, B. A.; James, T. L. MARDIGRAS-A Procedure for Matrix Analysis of Relaxation for Discerning Geometry of an Aqueous Structure. *J. Magn. Reson.* **1990**, *87* (3), 475–487. https://doi.org/10.1016/0022-2364(90)90305-S.

(181) James, T. L. Assessment of Quality of Derived Macromolecular Structures. *Methods Enzymol.* **1994**, *239* (C), 416–439. https://doi.org/10.1016/S0076-6879(94)39016-9.

(182) Lavery, R.; Sklenar, H. Defining the Structure of Irregular Nucleic Acids: Conventions and Principles. *J. Biomol. Struct. Dyn.* **1989**, *6* (4), 655–667.

https://doi.org/10.1080/07391102.1989.10507728.

(183) Mir, B. Studies on I-Motif Structures. Looking for Stable Structures at Neutral PH. *Master Univ. Química Orgànica. Univ. Barcelona.* **2014**.

(184) Solés, X. Studies on I-Motif Structures by NMR: Looking for Stable i-Motif Structures at Neutral PH. *Master Univ. Química Orgànica. Univ. Barcelona.* **2015**.

(185) Kan, L.-S.; Lin, W.-C.; Yadav, R. D.; Shih, J. H.; Chao, I. NMR Studies of the Tautomerism in Pseudoisocytidine. *Nucleosides and Nucleotides* **1999**, *18* (4–5), 1091–1093. https://doi.org/10.1080/15257779908041655.

(186) Parsch, U.; Engels, J. W. PH-Independent Triple-Helix Formation with 6-Oxocytidine as Cytidine Analogue. *Chem. - A Eur. J.* **2000**, *6* (13), 2409–2424. https://doi.org/10.1002/1521-3765(20000703)6:13<2409::AID-CHEM2409>3.0.CO;2-H.

(187) Ono, A.; Ts'o, P. O. P.; Kan, L. S. Triplex Formation of an Oligonucleotide Containing 2'-O-Methylpseudoisocytidine with a DNA Duplex at Neutral PH. *J. Org. Chem.* **1992**, *57* (11), 3225–3230. https://doi.org/10.1021/jo00037a048.

(188) Chin, T. M.; Lin, S. B.; Lee, S. Y.; Chang, M. L.; Cheng, A. Y. Y.; Chang, F. C.; Pasternack, L.; Huang, D. H.; Kan, L. S. "Paper-Clip" Type Triple Helix Formation by 5'-d-(TC)3T(a)(CT)3C(b)(AG)3 (a and b = 0-4) as a Function of Loop Size with and without the Pseudoisocytosine Base in the Hoogsteen Strand. *Biochemistry* **2000**, *39* (40), 12457–12464. https://doi.org/10.1021/bi0004201.

(189) Shahid, K. A.; Majumdar, A.; Alam, R.; Liu, S. T.; Kuan, J. Y.; Sui, X.; Cuenoud, B.; Glazer, P. M.; Miller, P. S.; Seidman, M. M. Targeted Cross-Linking of the Human β-Globin Gene in Living Cells Mediated by a Triple Helix Forming Oligonucleotide. *Biochemistry* **2006**, *45* (6), 1970–1978. https://doi.org/10.1021/bi0520986.

(190) Wright, E. P.; Lamparska, K.; Smith, S. S.; Waller, Z. A. E. Substitution of Cytosine with Guanylurea Decreases the Stability of I-Motif DNA. *Biochemistry* **2017**, *56* (36), 4879–4883. https://doi.org/10.1021/acs.biochem.7b00628.

(191) Escaja, N.; Viladoms, J.; Garavís, M.; Villasante, A.; Pedroso, E.; González, C. A Minimal I-Motif Stabilized by Minor Groove G:T:G:T Tetrads. *Nucleic Acids Res.* **2012**, *40* (22), 11737–11747. https://doi.org/10.1093/nar/gks911.

(192) Gallego, J.; Chou, S.-H.; Reid, B. R. Centromeric Pyrimidine Strands Fold into an Intercalated Motif by Forming a Double Hairpin with a Novel T:G:G:T Tetrad: Solution Structure of the d(TCCCGTTTCCA) Dimer. *J. Mol. Biol.* **1997**, *273* (4), 840–856.

https://doi.org/10.1006/JMBI.1997.1361.

(193) Šponer, J.; Mládek, A.; Šponer, J. E.; Svozil, D.; Zgarbová, M.; Banáš, P.; Jurečka, P.; Otyepka, M. The DNA and RNA Sugar-Phosphate Backbone Emerges as the Key Player. An Overview of Quantum-Chemical, Structural Biology and Simulation Studies. *Physical Chemistry Chemical Physics*. November 28, 2012, pp 15257–15277. https://doi.org/10.1039/c2cp41987d.

(194) Kang, C. H.; Berger, I.; Lockshin, C.; Ratliff, R.; Moyzis, R.; Rich, A. Crystal Structure of Intercalated Four-Stranded d(C3T) at 1.4 Å Resolution. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91* (24), 11636–11640. https://doi.org/10.1073/pnas.91.24.11636.

(195) Leroy, J. L.; Guéron, M. Solution Structures of the I-Motif Tetramers of d(TCC), d(5methylCCT) and d(T5methylCC): Novel NOE Connections between Amino Protons and Sugar Protons. *Structure* **1995**, *3* (1), 101–120. https://doi.org/10.1016/S0969-2126(01)00138-1.

(196) Phan, A. T.; Guéron, M.; Leroy, J. L. The Solution Structure and Internal Motions of a Fragment of the Cytidine-Rich Strand of the Human Telomere. *J. Mol. Biol.* **2000**, *299* (1), 123–144. https://doi.org/10.1006/jmbi.2000.3613.

(197) Wright, E. P.; Huppert, J. L.; Waller, Z. A. E. Identification of Multiple Genomic DNA Sequences Which Form I-Motif Structures at Neutral PH. *Nucleic Acids Res.* **2017**, *45* (6), 2951-29591. Wright, E. P., Huppert, J. L. & Waller. https://doi.org/10.1093/nar/gkx090.

(198) Kumar, N.; Nielsen, J. T.; Maiti, S.; Petersen, M. I-Motif Formation with Locked Nucleic Acid (LNA). *Angew. Chemie Int. Ed.* **2007**, *46* (48), 9220–9222. https://doi.org/10.1002/anie.200701667.

(199) Pasternak, A.; Wengel, J. Modulation of I-Motif Thermodynamic Stability by the Introduction of UNA (Unlocked Nucleic Acid) Monomers. *Bioorg. Med. Chem. Lett.* **2011**, *21* (2), 752–755. https://doi.org/10.1016/J.BMCL.2010.11.106.

(200) Mir, B.; Solés, X.; González, C.; Escaja, N. The Effect of the Neutral Cytidine Protonated Analogue Pseudoisocytidine on the Stability of I-Motif Structures. *Sci. Rep.* **2017**, *7* (1), 2772. https://doi.org/10.1038/s41598-017-02723-y.

(201) Tereshko, V.; Urpí, L.; Malinina, L.; Huynh-Dinh, T.; Subirana, J. A. Structure of the B-DNA Oligomers d(CGCTAGCG) and d(CGCTCTAGAGCG) in New Crystal Forms. *Biochemistry* **1996**, *35* (36), 11589–11595. https://doi.org/10.1021/bi961048l.

(202) Tereshko, V.; Subirana, J. A. Influence of Packing Interactions on the Average Conformation

of B-DNA in Crystalline Structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **1999**, *55* (4), 810–819. https://doi.org/10.1107/S0907444999000591.

(203) Subirana, J. A.; Abrescia, N. G. A. Extra-Helical Guanine Interactions in DNA. *Biophys. Chem.* **2000**, *86* (2–3), 179–189. https://doi.org/10.1016/S0301-4622(00)00168-X.

(204) Tippin, D. B.; Sundaralingam, M. Nine Polymorphic Crystal Structures of d(CCGGGCCCGG), d(CCGGGCCm5CGG), d(Cm5CGGGCCm5CGG) and d(CCGGGCC(Br)5CGG) in Three Different Conformations: Effects of Spermine Binding and Methylation on the Bending and Condensation of A-DNA. *J. Mol. Biol.* **1997**, *267* (5), 1171–1185. https://doi.org/10.1006/jmbi.1997.0945.

(205) Ramakrishnan, B.; Sundaralingam, M. High Resolution Crystal Structure of the A-DNA Decamer d(CCCGCCGGG). Novel Intermolecular Base-Paired G * (G·C) Triplets. *J. Mol. Biol.* **1993**, *231* (2), 431–444. https://doi.org/10.1006/jmbi.1993.1292.

(206) Leonard, G. A.; Zhang, S.; Peterson, M. R.; Harrop, S. J.; Helliwell, J. R.; Cruse, W. B.; Langlois d'Estaintot, B.; Kennard, O.; Brown, T.; Hunter, W. N. Self-Association of a DNA Loop Creates a Quadruplex: Crystal Structure of d(GCATGCT) at 1.8 å Resolution. *Structure* **1995**, *3* (4), 335–340. https://doi.org/10.1016/S0969-2126(01)00165-4.

(207) Salisbury, S. A.; Wilson, S. E.; Powell, H. R.; Kennard, O.; Lubini, P.; Sheldrick, G. M.; Escaja, N.; Alazzouzi, E.; Grandas, A.; Pedroso, E. The Bi-Loop, a New General Four-Stranded DNA Motif. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (11), 5515–5518. https://doi.org/10.1073/pnas.94.11.5515.

(208) Gonzalez, C.; Escaja, N.; Rico, M.; Pedroso, E. NMR Structure of Two Cyclic Oligonucleotides. A Monomer-Dimer Equilibrium between Dumbbell and Quadruplex Structures [3]. *Journal of the American Chemical Society*. American Chemical Society 1998, pp 2176–2177. https://doi.org/10.1021/ja973691g.

(209) Escaja, N.; Pedroso, E.; Rico, M.; González, C. Dimeric Solution Structure of Two Cyclic Octamers: Four-Stranded DNA Structures Stabilized by A:T:A:T and G:C:G:C Tetrads. *J. Am. Chem. Soc.* **2000**, *122* (51), 12732–12742. https://doi.org/10.1021/ja002778q.

(210) Escaja, N.; Gómez-Pinto, I.; Pedroso, E.; González, C. Four-Stranded DNA Structures Can Be Stabilized by Two Different Types of Minor Groove G:C:G:C Tetrads. *J. Am. Chem. Soc.* **2007**, *129* (7), 2004–2014. https://doi.org/10.1021/ja066172z.

(211) Viladoms, J.; Escaja, N.; Frieden, M.; Gómez-Pinto, I.; Pedroso, E.; González, C. Self-Association of Short DNA Loops through Minor Groove C:G:G:C Tetrads. *Nucleic Acids Res.* **2009**, *37* (10), 3264–3275. https://doi.org/10.1093/nar/gkp191.

(212) Viladoms, J.; Escaja, N.; Pedroso, E.; González, C. Self-Association of Cyclic Oligonucleotides through G:T:G:T Minor Groove Tetrads. *Bioorganic Med. Chem.* **2010**, *18* (11), 4067–4073. https://doi.org/10.1016/j.bmc.2010.04.018.

(213) Kettani, A.; Kumar, R. A.; Patel, D. J. Solution Structure of a DNA Quadruplex Containing the Fragile X Syndrome Triplet Repeat. *J. Mol. Biol.* **1995**, *254* (4), 638–656. https://doi.org/10.1006/jmbi.1995.0644.

(214) Kettani, A.; Bouaziz, S.; Gorin, A.; Zhao, H.; Jones, R. A.; Patel, D. J. Solution Structure of a Na Cation Stabilized DNA Quadruplex Containing G·G·G·G· and G·C·G·C· Tetrads Formed by G-G-G-C Repeats Observed in Adeno-Associated Viral DNA. *J. Mol. Biol.* **1998**, *282* (3), 619–636. https://doi.org/10.1006/jmbi.1998.2030.

(215) Da Silva, M. W. Association of DNA Quadruplexes through G:C:G:C Tetrads. Solution Structure of d(GCGGTGGAT). *Biochemistry* **2003**, *42* (49), 14356–14365. https://doi.org/10.1021/bi0355185.

(216) Parkinson, G. N.; Lee, M. P. H.; Neidle, S. Crystal Structure of Parallel Quadruplexes from Human Telomeric DNA. *Nature* **2002**, *417* (6891), 876–880. https://doi.org/10.1038/nature755.

(217) Jing, N.; Hogan, M. E. Structure-Activity of Tetrad-Forming Oligonucleotides as a Potent Anti-HIV Therapeutic Drug. *J. Biol. Chem.* **1998**, *273* (52), 34992–34999. https://doi.org/10.1074/jbc.273.52.34992.

(218) Bouaziz, S.; Kettani, A.; Patel, D. J. A K Cation-Induced Conformational Switch within a Loop Spanning Segment of a DNA Quadruplex Containing G-G-G-C Repeats. *J. Mol. Biol.* **1998**, *282* (3), 637–652. https://doi.org/10.1006/jmbi.1998.2031.

(219) Zhang, N.; Gorin, A.; Majumdar, A.; Kettani, A.; Chernichenko, N.; Skripkin, E.; Patel, D. J. Dimeric DNA Quadruplex Containing Major Groove-Aligned A·T·A·T and G·C·G·C Tetrads Stabilized by Inter-Subunit Watson-Crick A·T and G·C Pairs. *J. Mol. Biol.* **2001**, *312* (5), 1073–1088. https://doi.org/10.1006/jmbi.2001.5002.

(220) Gallego, J.; Chou, S. H.; Reid, B. R. *Centromeric Pyrimidine Strands Fold into an Intercalated Motif by Forming a Double Hairpin with a Novel T:G:G:T Tetrad: Solution Structure of the d(TCCCGTTTCCA) Dimer*; 1997; Vol. 273. https://doi.org/10.1006/jmbi.1997.1361.

(221) Viladoms, J.; Escaja, N.; Frieden, M.; Gómez-Pinto, I.; Pedroso, E.; González, C. Self-Association of Short DNA Loops through Minor Groove C:G:G:C Tetrads. *Nucleic Acids Res.* **2009**, *37* (10), 3264–3275. https://doi.org/10.1093/nar/gkp191.

(222) Šponer, J.; Jurečka, P.; Hobza, P. Accurate Interaction Energies of Hydrogen-Bonded Nucleic

Acid Base Pairs. *J. Am. Chem. Soc.* **2004**, *126* (32), 10142–10151. https://doi.org/10.1021/ja048436s.

(223) Serrat, G. Stabilization of I-Motif Structures. Converting Bimolecular i-Motif into Unimolecular. *Master Univ. Química Orgànica. Univ. Barcelona.* **2016**.

(224) Escaja, N.; Gómez-Pinto, I.; Viladoms, J.; Pedroso, E.; González, C. The Effect of Loop Residues in Four-Stranded Dimeric Structures Stabilized by Minor Groove Tetrads. *Org. Biomol. Chem.* **2013**, *11* (29), 4804. https://doi.org/10.1039/c3ob40741a.

(225) McLean, C. Y.; Bristor, D.; Hiller, M.; Clarke, S. L.; Schaar, B. T.; Lowe, C. B.; Wenger, A. M.; Bejerano, G. GREAT Improves Functional Interpretation of Cis-Regulatory Regions. *Nat. Biotechnol.* **2010**, *28* (5), 495–501. https://doi.org/10.1038/nbt.1630.

(226) Mir, B.; Serrano, I.; Buitrago, D.; Orozco, M.; Escaja, N.; González, C. Prevalent Sequences in the Human Genome Can Form Mini I-Motif Structures at Physiological PH. *J. Am. Chem. Soc.* **2017**, *139* (40), 13985–13988. https://doi.org/10.1021/jacs.7b07383.

(227) McCluskey, K.; Shaw, E.; Lafontaine, D. A.; Penedo, J. C. Single-Molecule Fluorescence of Nucleic Acids. *Methods Mol. Biol.* **2014**, *1076*, 759–791. https://doi.org/10.1007/978-1-62703-649-8_35.

(228) Mata, G.; Luedtke, N. W. Fluorescent Probe for Proton-Coupled DNA Folding Revealing Slow Exchange of *i* -Motif and Duplex Structures. *J. Am. Chem. Soc.* **2015**, *137* (2), 699–707. https://doi.org/10.1021/ja508741u.

(229) Zhang, S.; Sun, H.; Yang, D.; Liu, Y.; Zhang, X.; Chen, H.; Li, Q.; Guan, A.; Tang, Y. Evaluation of the Selectivity of G-Quadruplex Ligands in Living Cells with a Small Molecule Fluorescent Probe. *Anal. Chim. Acta X* **2019**, *2*, 100017. https://doi.org/10.1016/j.acax.2019.100017.

(230) Ward, D. C.; Reich, E.; Stryer, L. Fluorescence Studies of Nucleotides and Polynucleotides. I. Formycin, 2-Aminopurine Riboside, 2,6-Diaminopurine Riboside, and Their Derivatives. *J. Biol. Chem.* **1969**, *244* (5), 1228–1237.

(231) Rachofsky, E. L.; Osman, R.; Ross, J. B. A. Probing Structure and Dynamics of DNA with 2-Aminopurine: Effects of Local Environment on Fluorescence. *Biochemistry* **2001**, *40* (4), 946–956. https://doi.org/10.1021/bi001664o.

(232) Holz, B.; Klimasauskas, S.; Serva, S.; Weinhold, E. 2-Aminopurine as a Fluorescent Probe for DNA Base Flipping by Methyltransferases. *Nucleic Acids Res.* **1998**, *26* (4), 1076–1083. https://doi.org/10.1093/nar/26.4.1076.

(233) Neely, R. K.; Daujotyte, D.; Grazulis, S.; Magennis, S. W.; Dryden, D. T. F.; Klimašauskas, S.; Jones, A. C. Time-Resolved Fluorescence of 2-Aminopurine as a Probe of Base Flipping in M.HhaI - DNA Complexes. *Nucleic Acids Res.* **2005**, *33* (22), 6953–6960. https://doi.org/10.1093/nar/gki995.

(234) Hawkins, M. E.; Pfieiderer, W.; Mazumder, A.; Pommier, Y. G.; Balis, F. M. Incorporation of a Fluorescent Guanosine Analog into Olignoucleotides and Its Application to a Real Time Assay for the HIV-1 Integrase 3′-Processing Reaction. *Nucleic Acids Res.* **1995**, *23* (15), 2872–2880. https://doi.org/10.1093/nar/23.15.2872.

(235) Hawkins, M. E.; Pfleiderer, W.; Jungmann, O.; Balis, F. M. Synthesis and Fluorescence Characterization of Pteridine Adenosine Nucleoside Analogs for DNA Incorporation. *Anal. Biochem.* **2001**, *298* (2), 231–240. https://doi.org/10.1006/abio.2001.5399.

(236) Ren, R. X. F.; Chaudhuri, N. C.; Paris, P. L.; Rumney IV, S.; Kool, E. T. Naphthalene, Phenanthrene, and Pyrene as DNA Base Analogues: Synthesis, Structure, and Fluorescence in DNA. *J. Am. Chem. Soc.* **1996**, *118* (33), 7671–7678. https://doi.org/10.1021/ja9612763.

(237) Korshun, V. A.; Prokhorenko, I. A.; Gontarev, S. V.; Skorobogatyi, M. V.; Balakin, K. V.; Manasova, E. V.; Malakhov, A. D.; Berlin, Y. A. New Pyrene Derivatives for Fluorescent Labeling of Oligonucleotides. In *Nucleosides and Nucleotides*; Taylor & Francis Group, 1997; Vol. 16, pp 1461–1464. https://doi.org/10.1080/07328319708006206.

(238) Chan, K. M.; Kölmel, D. K.; Wang, S.; Kool, E. T. Color-Change Photoswitching of an Alkynylpyrene Excimer Dye. *Angew. Chemie - Int. Ed.* **2017**, *56* (23), 6497–6501. https://doi.org/10.1002/anie.201701235.

(239) Debnath, M.; Fatma, K.; Dash, J. Chemical Regulation of DNA I-Motifs for Nanobiotechnology and Therapeutics. *Angew. Chemie Int. Ed.* **2019**, *58* (10), 2942–2957. https://doi.org/10.1002/anie.201813288.

(240) Mergny, J.-L. Fluorescence Energy Transfer as a Probe for Tetraplex Formation: The i-Motif. **1999**. https://doi.org/10.1021/bi982208r.

(241) Börjesson, K.; Preus, S.; El-Sagheer, A. H.; Brown, T.; Albinsson, B.; Wilhelmsson, L. M. Nucleic Acid Base Analog FRET-Pair Facilitating Detailed Structural Measurements in Nucleic Acid Containing Systems. *J. Am. Chem. Soc.* **2009**, *131* (12), 4288–4293. https://doi.org/10.1021/ja806944w.

(242) Choi, J.; Majima, T. Reversible Conformational Switching of I-Motif DNA Studied by Fluorescence Spectroscopy. *Photochemistry and Photobiology*. 2013, pp 513–522. https://doi.org/10.1111/php.12042.

(243) Reilly, S. M.; Lyons, D. F.; Wingate, S. E.; Wright, R. T.; Correia, J. J.; Jameson, D. M.; Wadkins, R. M. Folding and Hydrodynamics of a DNA I-Motif from the c-MYC Promoter Determined by Fluorescent Cytidine Analogs. *Biophys. J.* **2014**, *107* (7), 1703–1711. https://doi.org/10.1016/j.bpj.2014.08.014.

(244) Benabou, S.; Ruckebusch, C.; Sliwa, M.; Aviñó, A.; Eritja, R.; Gargallo, R.; de Juan, A. Study of Conformational Transitions of I-Motif DNA Using Time-Resolved Fluorescence and Multivariate Analysis Methods. *Nucleic Acids Res.* **2019**, *47* (13), 6590–6605. https://doi.org/10.1093/nar/gkz522.

(245) Lin, K. Y.; Jones, R. J.; Matteucci, M. Tricyclic 2'-Deoxycytidine Analogs: Syntheses and Incorporation into Oligodeoxynucleotides Which Have Enhanced Binding to Complementary RNA. *J. Am. Chem. Soc.* **1995**, *117* (13), 3873–3874. https://doi.org/10.1021/ja00118a026.

(246) Marcus Wilhelmsson, L.; Holmén, A.; Lincoln, P.; Nielsen, P. E.; Nordén, B. A Highly Fluorescent DNA Base Analogue That Forms Watson-Crick Base Pairs with Guanine. *Journal of the American Chemical Society*. American Chemical Society 2001, pp 2434–2435. https://doi.org/10.1021/ja0025797.

(247) Bielecka, P.; Juskowiak, B. Fluorescent Sensor for PH Monitoring Based on an I-Motif- – Switching Aptamer Containing a Tricyclic Cytosine Analogue (TC). *Molecules* **2015**, *20* (10), 18511–18525. https://doi.org/10.3390/molecules201018511.

(248) Sandin, P.; Börjesson, K.; Li, H.; Mårtensson, J.; Brown, T.; Wilhelmsson, L. M.; Albinsson, B. Characterization and Use of an Unprecedentedly Bright and Structurally Non-Perturbing Fluorescent DNA Base Analogue. *Nucleic Acids Res.* **2008**, *36* (1), 157–167. https://doi.org/10.1093/nar/gkm1006.

(249) Engman, K. C.; Sandin, P.; Osborne, S.; Brown, T.; Billeter, M.; Lincoln, P.; Nordén, B.; Albinsson, B.; Wilhelmsson, L. M. DNA Adopts Normal B-Form upon Incorporation of Highly Fluorescent DNA Base Analogue TC: NMR Structure and UV-Vis Spectroscopy Characterization. *Nucleic Acids Res.* **2004**, *32* (17), 5087–5095. https://doi.org/10.1093/nar/gkh844.

(250) Preus, S.; Börjesson, K.; Kilså, K.; Albinsson, B.; Wilhelmsson, L. M. Characterization of Nucleobase Analogue FRET Acceptor TCnitro. *J. Phys. Chem. B* **2010**, *114* (2), 1050–1056. https://doi.org/10.1021/jp909471b.

(251) Beaucage, S. L.; Caruthers, M. H. Deoxynucleoside Phosphoramidites-A New Class of Key Intermediates for Deoxypolynucleotide Synthesis. *Tetrahedron Lett.* **1981**, *22* (20), 1859–1862. https://doi.org/10.1016/S0040-4039(01)90461-7.

(252) Matteucci, M. D.; Caruthers, M. H. Synthesis of Deoxyoligonucleotides on a Polymer Support. *J. Am. Chem. Soc.* **1981**, *103* (11), 3185–3191. https://doi.org/10.1021/ja00401a041.

(253) Alazzouzi, E. M.; Escaja, N.; Grandas, A.; Pedroso, E. A Straightforward Solid-Phase Synthesis of Cyclic Oligodeoxyribonucleotides. *Angew. Chemie (International Ed. English)* **1997**, *36* (13–14), 1506–1508. https://doi.org/10.1002/anie.199715061.

(254) Tataurov, A. V.; You, Y.; Owczarzy, R. Predicting Ultraviolet Spectrum of Single Stranded and Double Stranded Deoxyribonucleic Acids. *Biophys. Chem.* **2008**, *133* (1–3), 66–70. https://doi.org/10.1016/j.bpc.2007.12.004.

(255) Sandin, P.; Bö Rjesson, K.; Li, H.; Må Rtensson, J.; Brown, T.; Wilhelmsson, L. M.; Albinsson, B. Characterization and Use of an Unprecedentedly Bright and Structurally Non-Perturbing Fluorescent DNA Base Analogue. *Nucleic Acids Res.* **2008**, *36* (1), 157–167. https://doi.org/10.1093/nar/gkm1006.

(256) Lee, W.; Tonelli, M.; Markley, J. L. NMRFAM-SPARKY: Enhanced Software for Biomolecular NMR Spectroscopy. *Bioinformatics* **2015**, *31* (8), 1325–1327. https://doi.org/10.1093/bioinformatics/btu830.

RESUM EN CATALÀ

Les estructures i-motif són estructures no canòniques de DNA que foren descrites l'any 1993 com la primera associació intercalada entre cadenes de desoxiribonucleòtids. La baixa estabilitat a pH neutre de les primeres seqüències estudiades, pel fet de presentar com a unitat estructural bàsica parells hemiprotonats $C:C^+$, va fer dubtar inicialment a la comunitat científica de la seva possible rellevància *in vivo*. Contràriament, una extensa quantitat d'aplicacions nanotecnològiques es van desenvolupar basades en estructures i-motif i les interessants transicions estructurals motivades per modificacions de pH que aquestes presenten.

La recerca realitzada en els darrers anys en el camp de les estructures i-motif, però ha posat de relleu la potencial implicació d'aquestes estructures en processos biològics. Centrades en seqüències riques en citosina complementàries a aquelles conegudes per formar estructures de quàdruplexs de guanina (certament les estructures no canòniques de DNA més estudiades), nombroses publicacions proposen la formació d'estructures i-motif en zones promotores d'oncogens, en telòmers i en centròmers. Fins i tot, alguns estudis indaguen en la potencial funció reguladora d'aquestes estructures aportant evidències de petites molècules capaces de modular aquesta funció. L'exponencial interès a nivell biològic de les estructures i-motif culminà l'any 2018 amb dues publicacions assenyalant l'observació d'estructures i-motif en medi cel·lular mitjançant dues aproximacions totalment diferents.

En aquest context, la recerca recent desenvolupada en el nostre grup de recerca va identificar una seqüència curta, sense citosines contigües, capaç de formar estructures i-motif dimèriques observables a pH neutre. L'estructura en solució d'aquest i-motif conté únicament dos parells hemiprotonats $C:C^+$. Apilades sobre aquests parells, dues tètrades G:T:G:T de solc menor prèviament descrites en una altra estructura i-motif, actuen com a tapa , conferint-li una elevada estabilitat. Aquests mini i-motifs amplien el ventall de seqüències capaces de formar estructures i-motif, posant de manifest que no estan limitades a seqüències riques en citosina. A més, evidencien que d'altres elements estructurals, com les tètrades de solc menor poden contribuir de forma important a la seva estabilitat.

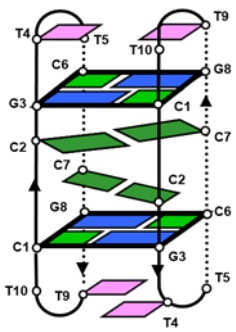


Figura 1. Representació esquemàtica d'una estructura i-motif dimèrica formada per oligonucleòtids cíclics.

Els objectius plantejats per aquesta tesi es varen centrar doncs en l'estudi d'estructures tipus mini i-motif bàsicament orientats en dues línies de treball. D'una banda, un dels objectius generals està enfocat en el disseny i estudi estructural detallat de diferents seqüències que permetessin determinar la compatibilitat i l'efecte estabilitzant en i-motifs d'altres tipus de tètrades de solc menor observades en altres formes no canòniques. A la Taula 1 s'han agrupat les tres famílies principals de seqüències estudiades en aquesta tesi

NOM	SEQÜÈNCIA	TÈTRADES
M	d(TCGTTTCGT)	G:T:G:T
cM	d<pTCGTTTCGTT>	
MM4	d(TCGTTTCGT-TTTT-TCGTTTCGT)	
L	d(TCGTTCCGT)	2x G:C:G:T o 1x G:T:G:T + 1x G:C:G:C
cL	d<pTCGTTCCGTT>	
LL4	d(TCGTTCCGT-TTTT-TCGTTCCGT)	
N	d(CCGTTCCGT)	G:C:G:C
cN	d<pCCGTTCCGTT>	
NN4	d(CCGTTCCGT-TTTT-CCGTTCCGT)	

Taula 1. Seqüències no modificades més estudiades en aquesta tesi i les possibles tètrades que poden formar donat el plegament en estructures mini i-motif.

El segon gran bloc d'objectius plantejats per a aquesta tesi, es centra en la incorporació estratègica en algunes de les seqüències presentades a la Taula 1 de diferents anàlegs de citidina. D'una banda, s'ha explorat la incorporació d'un anàleg neutre de citidina protonada, la pseudoisocitidina (psC) amb la finalitat d'estudiar la viabilitat d'obtenir estructures i-motif independents de pH. D'altra banda, s'ha estudiat la substitució de residus específics de citidina per un anàleg fluorescent derivat de fenoxazina (tC^o), amb l'objectiu d'utilitzar-lo com a sonda interna per a monitoritzar canvis estructurals a les estructures i-motif. En tots dos casos, es manté el patró de formació de ponts d'hidrogen de la citosina i s'espera que la seva incorporació no afecti de forma dràstica a la formació dels i-motifs. A la Figura 2 es mostra l'estructura dels parells que formarien aquest dos anàlegs amb citidina.

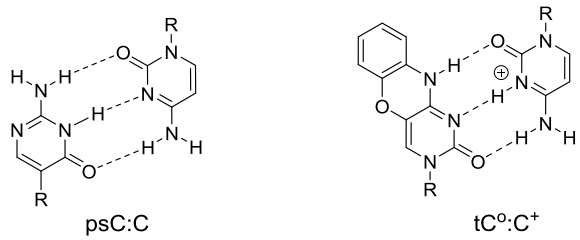


Figura 2. Anàlegs de citosina estudiats en aquesta tesi formant respectius parells psC:C i $tC^{o}:C^{+}$.

El primer capítol de la tesi està centrat en la incorporació de residus de psC en posicions especifiques de **M** i **MC2**. En tots dos casos es formen estructures dimèriques. Els resultats indiquen una gran dependència de l'estabilitat de les estructures resultants amb la posició que ocupa l'anàleg neutre a l'estructura. En termes generals, els estudis d'estabilitat monitoritzats per RMN i DC realitzats per les diferents seqüències estudiades mostren una clara desestabilització de les estructures. Únicament en el cas de la seqüència MC2_psC(9), que conté quatre parells hemiprotonats C:C+ i els residus de psC es troben als parells terminals, s'observa un increment d'estabilitat que permet la visualització d'aquesta estructura perfectament plegada a pH neutre l'estructura en solució d'aquesta seqüència, que es mostra a la Figura 3, fou calculada a partir de l'assignació completa d'espectres NOESY a pH 7. El plegament de l'estructura està afavorit per la formació de dos parells $C:C^{+}$ al centre de l'estructura, dos parells psC:C flanquejant el centre hemiprotonat de l'estructura i dos parells G:G apilats sobre els psC:C als extrems del motiu.

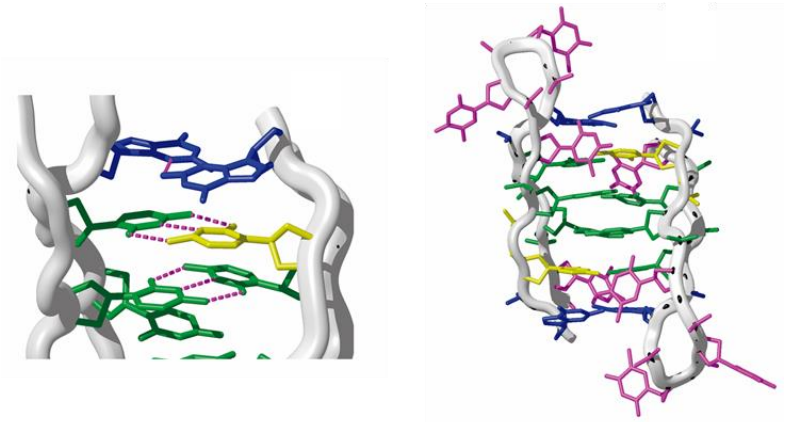


Figura 3. Estructura mitjana en solució calculada per **MC2_psC(9)** a pH 7. Citosines mostrades en verd, psC en groc, guanines en blau i timines en fúcsia.

Per confirmar la generalitat dels resultats obtinguts per les seqüències anteriors, la psC fou incorporada també en una seqüència telomèrica que es coneix que forma un i-motif unimolecular: d(CCCTAACCCTAACCCTAACCC). Efectivament, els resultats d'estabilitat obtinguts per aquestes seqüències confirmen que la psC, incorporada en posicions centrals de l'estructura, provoca una forta desestabilització del motiu. En canvi, la seva presencia en posicions situades als extrems de l'estructura (com en el cas de **MC2_psC(9)**) si que es tolerada per la formació del motiu i pot fins i tot conferir una certa estabilitat extra a l'estructura.

En general, aquests resultats evidencien la gran importància que tenen els parells hemiprotonats en la formació d'estructures i-motif, molt probablement relacionada amb l'alleujament de repulsions electroestàtiques entre cadenes a través del solc menor. A més, relacionat amb la formació de tètrades de solc menor, sembla que aquestes afavoreixen interaccions de tipus catió-pi i que requereixen la presència de parells hemiprotonats per la seva formació.

Dos dels capítols de la tesi estan centrats en l'estudi de la compatibilitat de tètrades de solc menor amb estructures i.motif i l'efecte que poden tenir a la seva estabilitat. Amb aquest objectiu, s'han estudiat diverses seqüències que formen i-motifs, tant dimèrics com monomèrics que, un cop plegades donen lloc als diferents tipus de tètrades. Una de les primeres conclusions que s'ha tret d'aquests estudis és la necessitat de que les tètrades siguin desplaçades, és a dir, que els parells WC que formen les tètrades no estiguin enfrontats directament (Figura 4). Així doncs, bàsicament els estudis detallats s'han centrat en seqüències potencialment capaces de formar estructures i-motif estabilitzades per tètrades G:T:G:T, G:C:G:C i G:C:G:T desplaçades.

Les construccions dimèriques, tant a partir de oligonucleòtids lineals com cíclics (**M**, **L**, **N**, **cM**, **cL** i **cN**), tot i presentar baixa estabilitat a pH neutre, permeten extreure vàries conclusions clau per comprendre millor les dinàmiques de les estructures mini i-motif:

- Per estructures en les que es formen el mateix tipus de tètrades independentment de la topologia (seqüències **M** i **N**), s'observa un equilibri aproximadament 1:1 entre les topologies cap-cap i cap-cua.

- Per la seqüència **L**, per la que la topologia condiciona el tipus de tètrada formada, només s'observa la formació de l'estructura cap-cua, amb formació de tètrades mixtes G:C:G:T, indicant una marcada preferència per aquesta tipus de tètrada

- La seqüència N, mostra un equilibri entre diferents estructures i-motif dependent de pH en les que hi ha protonació de citosines que formen les tètrades a pH neutre.

- Les estructures en les que es formen tètrades G:C:G:C o G:C:G:T són marcadament més estables que les que contenen tètrades G:T:G:T.

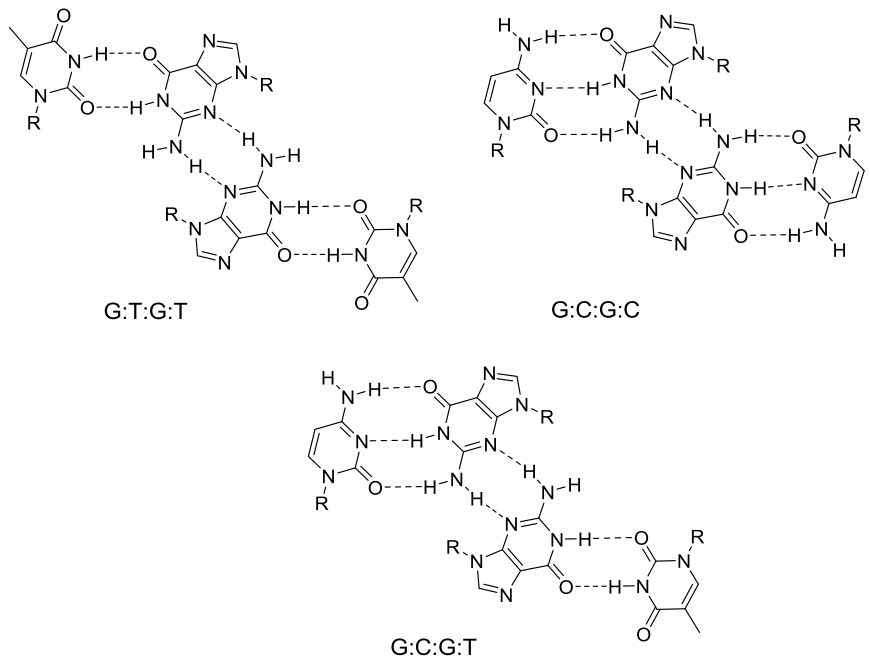


Figura 4. Tètrades de solc menor estudiades en aquesta tesi.

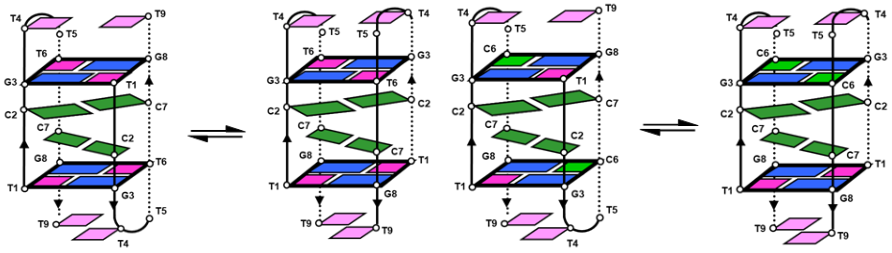


Figura 5. Diferents topologies dels dímers M2 (esquerra) i L2 (dreta).

El disseny de seqüències repetitives (**MM**, **LL** i **NN**) capaces de formar mini i-motif unimoleculars representa un dels acompliments més significatius d'aquesta tesi. La gran estabilitat mostrada per

aquestes estructures enfront de la desnaturalització per pH (pH_T = 8) és totalment inusitada per estructures i-motif. En aquest sentit, la unimolecularitat de les estructures i l'efecte de les tètrades de solc menor són clau. Per aquestes seqüències, reforçant els resultats trobats per les construccions dimèriques, es pot definir un rang d'estabilització proporcionat per les diferents tètrades: G:C:G:T ≥ G:C:G:C >>> G:T:G:T.

La caracterització de les estructures mitjançant experiments de RMN permeté comprovar que per les tres seqüències que foren estudiades (**MM4**, **LL4** i **NN4**), els principals trets estructurals són compartits per les tres estructures: topologia cap-cua i dos parells hemiprotonats $C:C^+$ al centre de l'estructura totalment apilats sota els parells de bases que conformen les tètrades.

L'efecte del linker central de timines que connecta les dues repeticions de la seqüència fou estudiat en detall. Mitjançant corbes de fusió de seqüències derivades de **LL4** amb diferent nombre de timines, es pot concloure que la llargària d'aquest linker no té cap efecte significatiu sobre la formació del motiu, però si sobre l'estabilitat tèrmica d'aquests, essent **LL3** i **LL4** les seqüències que exhibeixen una estabilitat més gran. D'altra banda, s'ha pogut confirmar que la topologia preferida d'aquestes estructures es cap-cua probablement degut a que el linker de timines ocupa el solc major de les estructures.

La Figura 6 mostra l'estructura en solució calculada per **LL3** a pH 7 mitjançant càlculs de dinàmica molecular a partir de les restriccions experimentals obtingudes a partir dels espectres de RMN. Com es pot observar, es tracta d'una estructura força compacta que proporciona una elevada protecció dels parells hemiprotonats.

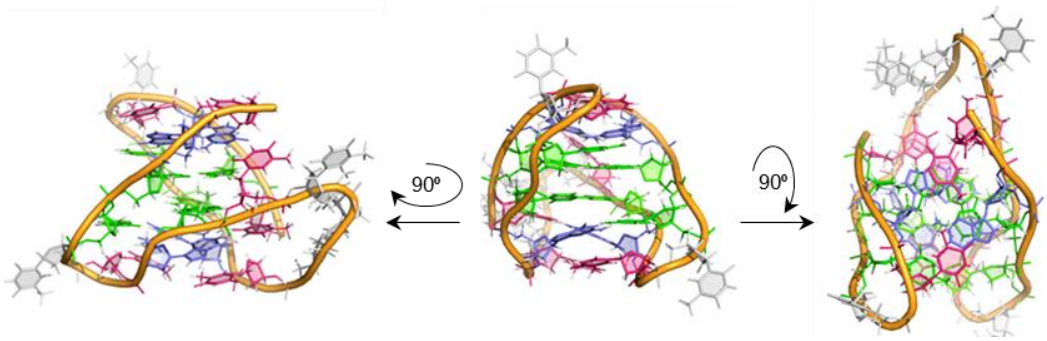


Figura 6. Diferents vistes de l'estructura en solució de **LL3** a pH 7. Citosines mostrades en verd, guanines en blau i timines estructurades en fúcsia i timines exposades en gris.

De forma interessant, la seqüència **NN4** exhibeix el mateix comportament front el pH que el seu anàleg curt N, que implica l'aparició d'una espècie i-motif alternativa a pH àcid. Aquesta estructura àcida alternativa fou caracteritzada per RMN i consisteix en un apilament de quatre pisos de parells $C:C^+$ amb dues tètrades G:T:G:T apilades als extrems de l'estructura. La formació d'aquesta estructura s'explica a partir de la protonació de les citosines que formen parells G:C en l'espècie neutra, re col·locant-se al centre de l'estructura per mantenir l'ordre d'apilament. A més, la nova estructura, inclou timines del loop central per la formació de les tètrades G:T:G:T.Motivats pels sorprenents resultats d'estabilitat obtinguts per aquestes estructures monomèriques, i després de comprovar que seqüències més llargues de fins a sis repeticions de la seqüència **L** també són capaces d'adoptar estructures mini i-motif es va determinar una seqüència consens per analitzar la seva presència en el genoma humà: d(YCG(XX)YCG(Xn)YCG(XX)YCG). Els resultats d'aquests estudis bioinformàtics revelen que aquesta seqüència es troba estadísticament sobreexpressada al genoma humà. A més, la seva localització no és arbitrària, sinó que es centra en regions properes a inici de transcripció de gens associats a funcions reguladores i de desenvolupament. Aquests resultats són molt prometedors a l'hora de considerar la possible rellevància biològica d'aquestes seqüències.

Finalment, s'ha volgut explorar l'eficiència de l'anàleg fluorescent de citosina tC^o per a la caracterització de transicions estructurals dins d'estructures i-motif incorporant aquest residu en posicions especifiques de la seqüència **NN4**. Aquesta seqüència és especialment interessant perquè té residus de citosina que, un cop formada l'estructura, es troben en ambients químics totalment diferents: bé formant un parell $tC^o:C^+$ (**NN4_tC^o(2)**) o bé un parell G:tCo de tipus WC en el context d'una tètrada G:C:G:C (**NN4_tC^o(6)**).

La caracterització d'ambdues seqüències revelà la formació d'estructures i-motif de forma anàloga a la seqüència original **NN4**, exhibint també un equilibri estructural dependent de pH entre dues formes i-motif: la forma "àcida" de quatre pisos de $C:C^+$ i la forma "neutra" mini i-motif. En termes d'estabilitat, **NN4_tC^o(2)** exhibeix una major resistència a la desnaturalització tèrmica de quasi bé +10 ºC comparada amb els seus **NN**-anàlegs. A més, segons mostren els experiments de RMN, la forma neutra és encara present a valors de pH àcid tot i coexistir amb la forma àcida. Aquests resultats s'expliquen a partir de l'estructura en solució calculada per **NN4_tC^o(2)** a pH 7. L'estructura, mostra com el residu tC^o2, a banda de formar perfectament un parell hemiprotonat es troba totalment apilat sota un parell G:C, compta amb interaccions favorables d'apilament extra pel seu sistema aromàtic més extens. Els detalls dels principals trets estructural d'aquesta construcció es mostren en la Figura 7.

Els estudis de DC i fluorescència permeteren determinar que la forma neutra de **NN4_tC^o(2)** és majoritària en el rang de pH 5.4 – 8.3, significativament més gran que el trobat per les altres estructures. A més, dels estudis de fluorescència, s'observa com l'extinció de la senyal de fluorescència és total

quan el parell tC^{o} es troba hemiprotonat (Figura 8), fet que permet identificar l'estat d'hibridació d'aquest residu, fent-lo molt útil per la caracterització de canvis subtils dins l'estructura.

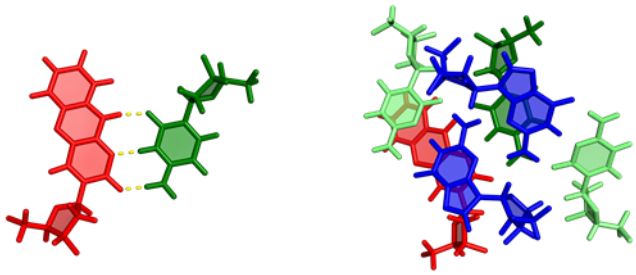


Figura 7. Detalls estructurals de l'estructura en solució de **NN4_tC^{o}(2)** a pH 7**.** Citosines hemiprotonades mostrades en verd fosc, en verd clar les que formen parells G:C, guanines en blau i tC^{o} en vermell.

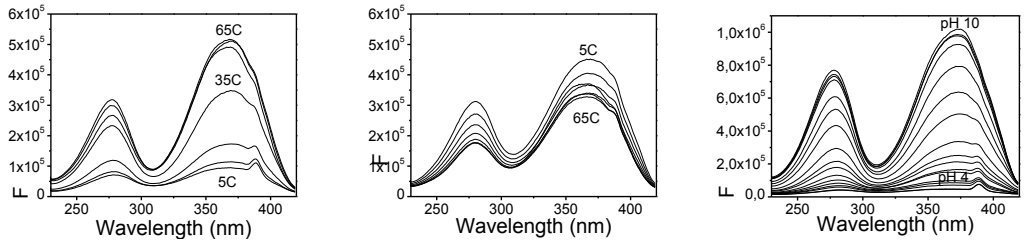


Figura 8. Espectres superposats d'excitació per la seqüència **NN4_tC^{o}(2)** a pH 7 i 5 ºC mostrant l'extinció de la senyal a pH àcid i fins que no es desnaturalitza l'estructura (dreta). Espectres superposats d'excitació per la seqüència **NN4_tC^{o}(2)** a diferent temperatura en (centre) i sense (esquerra) presencia de la seva seqüència complementaria.

Per completar els estudis de fluorescència, s'explorà la visualització d'estructures mini i-motif en cèl·lules mitjançant microscòpia confocal fluorescent. Amb aquest objectiu, prèviament es va comprovar *in vitro* que en presència de la seva seqüència complementària, la hibridació del dúplex no afecta la senyal de fluorescència emesa pel tC^{o}. Un cop confirmat, cèl·lules HeLa foren transfectades amb la seqüència **NN4_tC^{o}(2)** i es procedí a visualitzar-la al microscopi en medi fisiològic, contrastada amb una seqüència control negatiu incapaç de formar estructures i-motif. Comparada qualitativament, la senyal emesa per **NN4_tC^{o}(2)** és 50% menys intensa que la seqüència control, fet coherent amb l'estabilitat tèrmica determinada per **NN4_tC^{o}(2)** a pH 7 (38.8 ºC). A continuació, es monitoritzà

exitosament el plegament total de la seqüència, reflectit en la total desaparició d'emissió fluorescent, amb addicions progressives d'àcid cítric. Aquests notables resultats prometen molt en el sentit d'utilitzar el tC^{o} com a nucleobase fluorescent per a la visualització indirecta d'estructures i-motif *in vivo*.

APPENDIX 1

RESIDUE	H1/H3$^+$	H42/H22	H41/H21	H6/H8	H5/Me	H1'	H2'	H2"
T1	n.o.	-	-	7.76	1.86	6.42	2.52	2.70
C2	15.47	8.74	7.43	7.51	6.36		1.13	2.23
G3	10.75	8.63	5.77	8.28	-	5.93	2.93	2.69
T4	n.o.	-	-	7.62	1.72	6.12	2.02	2.31
T5	n.o.	-	-	7.82	1.92	6.47	2.29	2.55
C6	-	8.57	7.40	7.88	6.10	6.28	2.43	
C7	15.47	10.49	8.13	7.41	5.95	6.22	0.75	2.10
G8	13.74	8.04	7.61	8.27	-	6.00	3.03	2.64
T9	n.o.	-	-	7.62	1.72	5.93	2.12	

Table App 1. Chemical shifts of head-to-tail structure of **L** at pH 7 and 5 ºC. na: not assigned. no: not observed

RESIDUE	H1/H3$^+$	H42/H22	H41/H21	H6/H8	H5/Me	H1'	H2'	H2"
T1	11.58	-	-	7.73	1.74	6.39	2.42	2.60
C2	15.45	8.64	7.33	7.46	6.28	6.34	1.03	2.12
G3	10.64	8.52	7.53	8.27	-	5.89	2.82	2.53
T4	10.51	-	-	7.59	1.62	6.09	1.92	2.21
T5	n.o.	-	-	7.79	1.81	6.43	2.19	2.45
C6	-	8.51	7.58	7.87	6.02	6.26	2.34	
C7	15.45	10.40	8.07	7.37	5.87	6.19	0.65	2.00
G8	13.55	7.78	7.63	8.25	-	5.96	2.92	2.57
T9	10.51	-	-	7.59	1.62	5.87	1.99	

Table App 2. Chemical shifts of head-to-tail structure of **L** at pH 5 and 5 ºC. na: not assigned. no: not observed

RESIDUE	H1/H3$^+$	H42/H22	H41/H21	H6/H8	H5/Me	H1'	H2'	H2"
T1	11.53							
C2				7.40		6.35	0.81	1.94
G3				8.27				
T4								
T5								
C6				7.96				
C7	15.32	9.52	7.64	7.43	6.31	6.14	0.95	2.26
G8	10.61	8.61	6.07	8.30	-	5.78	2.81	2.48
T9								

Table App 3. Chemical shifts of head-to-head structure of **L** at pH 5 and 5 ºC. na: not assigned. no: not observed

RESIDUE	H1/H3$^+$	H42/H22	H41/H21	H6/H8	H5/Me	H1'	H2'	H2"	H3'	H4'	H5'/H5"
T1	11.82	-	-	7.73	1.89	6.43	2.48	2.63	4.98	4.39	n.a.
C2	15.59	9.94	7.47	7.53	6.41	6.36	1.12	2.14	4.81	4.41	n.a.
G3	11.04	8.65	5.64	8.34	-	5.96	2.99	2.66	5.16	4.71	3.91/4.12
T4	10.55	-	-	7.66	1.80	6.14	2.03	2.31	4.79		n.a.
T5	n.o.	-	-	7.84	1.93	6.48	2.31	2.56	4.66	4.58	n.a.
C6	-	8.61	7.49	7.91	6.15	6.29	2.45		4.81	4.36	n.a.
C7	15.59	10.33	8.06	7.38	5.93	6.23	0.77	2.10	4.72	4.41	n.a.
G8	13.78	8.10	7.73	8.31	-	6.01	3.03	2.66	5.15	4.65	3.96/4.11
T9	10.50	-	-	7.62	1.73	5.98	2.02	2.32	4.82	4.34	n.a.
T10	n.o.	-	-	7.73	1.91	6.38	2.32	2.59	4.72	4.33	n.a.

Table App 4. Chemical shifts of **cL** at pH 6 and 5 ºC. na: not assigned. no: not observed

RESIDUE	H1/H3$^+$	H42/H22	H41/H21	H6/H8	H5/Me	H1'	H2'	H2"
C1	-	7.99	7.26	7.90	6.05	6.33	2.42	2.60
C2	15.43	9.29	8.06	7.49	5.93	6.24	1.07	
G3	11.82	8.28	6.52	8.22	-	5.97	2.95	
T4	n.o.	-	-	7.63	1.74	n.a.	n.a.	
T5	n.o.	-	-	n.a.	n.a.	n.a.	n.a.	
C6	-	8.57	7.39	7.91	6.12	6.31	2.45	
C7	15.43	9.96	7.91	7.40	5.88	6.19	0.80	2.09
G8	13.54	7.95	7.64	8.24	-	6.02	3.01	2.68
T9	n.o.	-	-	7.64	1.75	n.a.	n.a.	n.a.

Table App 5. Chemical shifts of head-to-tail structure of **N** at pH 7 and 5 ºC. na: not assigned. no: not observed.

RESIDUE	H1/H3$^+$	H42/H22	H41/H21	H6/H8	H5/Me	H1'	H2'	H2"
C1	-	n.a.	n.a.	7.86	6.04	6.31	2.44	2.52
C2	15.53	9.70	7.76	7.45	5.85	6.20	0.93	2.08
G3	13.08	7.63		8.21	-	6.04	3.00	2.69
T4	n.o.	-	-	n.a.	1.75	n.a.	n.a.	n.a.
T5	n.o.	-	-	n.a.	n.a.	n.a.	n.a.	n.a.
C6	-	8.59	7.42	7.96	6.11	6.36	2.41	2.52
C7	15.33	9.59	8.03	7.46	5.89	6.16	0.98	2.16
G8	12.41	8.06	7.29	8.23	-	5.98	2.95	2.65
T9	n.o.	-	-	n.a.	1.74	n.a.	n.a.	n.a.

Table App 6. Chemical shifts of head-to-head structure of **N** at pH 7 and 5 ºC. na: not assigned. no: not observed.

RESIDUE	H1/H3$^+$	H42/H22	H41/H21	H6/H8	H5/Me	H1'	H2'	H2"
C1	15.70	9.45	8.56	8.00	8.59	6.56	2.12	2.55
C2	15.27	10.05	8.90	7.57	5.80	6.15	1.52	2.39
G3	10.64	8.69	6.18	8.26	-	5.88	2.89	2.63
T4	n.o.	-	-	n.a.	1.72	n.a.	n.a.	n.a.
T5	11.46	-	-	n.a.	n.a.	n.a.	n.a.	n.a.
C6	15.70	9.39	8.32	7.95	6.13	6.53	1.94	2.52
C7	15.27	9.00	7.60	7.51	6.01		1.52	2.32
G8	9.75	8.47	6.59	8.19	1.74-	5.93	2.07	2.58
T9	n.o.	-	-	n.a.	n.a.	n.a.	n.a.	n.a.

Table App 7. Chemical shifts of head-to-head structure of **N** at pH 5 and 5 ºC. na: not assigned. no: not observed

RESIDUE	H1/H3$^+$	H42/H22	H41/H21	H6/H8	H5/Me	H1'	H2'	H2"	H3'	H4'	H5'/H5"
C1	-	8.69	7.39	7.92	6.11	6.39	2.43	2.53	4.88	4.35	n.a.
C2	15.45	9.68	7.89	7.47	5.84	6.37	1.01	2.16	4.61	4.33	4.02/4.16
G3	13.22	7.83	7.66	8.19	-	6.08	3.01	2.69	5.07	4.62	3.98/8.20
T4	n.o.	-	-	7.62	1.77	6.11	2.02	2.37	4.79	3.85	
T5	n.o.	-	-	7.83	1.94	6.46	2.24	2.57	4.67	4.53	n.a.

Table App 8. Chemical shifts of head-to-head structure of **cN** at pH 7 and 5 ºC. na: not assigned. no: not observed

RESIDUE	H1/H3$^+$	H42/H22	H41/H21	H6/H8	H5/Me	H1'	H2'	H2"	H3'	H4'	H5'/H5"
C1	-	8.74	n.o.	8.05	6.25	6.40	2.44	2.54	4.91	4.42	4.17
C2	15.47	9.70	7.83	7.47	6.04	6.25	0.97	2.1	4.73	4.43	4.10
G3	n.o.	n.o.	n.o.	8.28	-	6.01	3.03	2.70	5.10	4.63	3.90/4.11
T4	n.o.	-	-	7.62	1.77	6.12	2.06	2.37	4.83	3.93	
T5	n.o.	-	-	7.81	1.94	6.46	2.27	2.57	4.65	4.53	3.90/4.04

Table App 9. Chemical shifts of head-to-head structure of **cN** at pH 5 and 5 ºC. na: not assigned. no: not observed

RESIDUE	H1/H3$^+$	H42/H22	H41/H21	H6/H8	H5/Me	H1'	H2'	H2"
^{m}C2/^{m}C15	15.39/15.37	9.20	7.62/7.42	7.45/7.41	1.91/1.96	6.31	1.10/1.17	2.16/2.22
G3/G16	11.15/11.08	8.56	5.81/5.59	8.19/8.23	-	5.91	2.94	2.67
T4/T9/T17/T22	no	-	-	7.64	1.76	na	na	na
C6/C19	-	8.58	7.39	7.86	6.08	6.28	2.44	
C7/C20	15.39/15.37	10.35	8.11	7.35	5.91	6.20	0.73	2.15
G8/G21	13.76	7.98	7.66	8.29	-	6.02	3.04	2.71

Table App 10. Chemical shifts of the protons in the structure of **LL4_mC(2,15)** at pH 7 and 5 ºC. T1, T10, T11, T12, T13 and T14 could not be unambiguously assigned. na: not assigned. no: not observed.

RESIDUE	H1/H3$^+$	H42/H22	H41/H21	H6/H8	H5/Me	H1'	H2'	H2"
T1	n. o.	-	-	7.80	1.94	6.39	2.43	2.59
^{m}C2	15.40	9.55	7.82	7.45	2.01	6.32	1.20	2.20
G3	11.13	8.59	5.82	8.24	-	5.92	2.94	2.67
T4	n. o.	-	-	7.63	1.74	6.13	2.03	2.30
T5/T18	n. o.	-	-	7.84	1.93	6.48	2.32	2.54
C6	-	8.57	7.40	7.87	6.09	6.28	2.45	
C7	15.44	10.45	8.16	7.36	5.92	6.20	0.73	2.09
G8	13.71	7.99	7.70	8.29	-	5.99	3.02	2.69
T9	n. o.	-	-	7.59	1.72	5.85	2.15	2.38
T10	n. o.	-	-	7.66	1.76	6.17	2.29	2.45
T14	n. o.	-	-	7.79	1.89	6.85	2.49	2.63
C15	15.44	8.71	7.53	7.58	6.29	6.33	1.05	2.19
G16	10.66	8.62	5.67	8.51	-	5.94	2.94	2.66
T17	n. o.	-	-	7.63	1.79	6.13	2.03	2.30
C19	-	8.58	7.41	7.87	6.09	6.28	2.45	
C20	15.40	10.07	7.91	7.35	5.89	6.20	1.37	2.17
G21	13.78	8.00	6.62	8.28	-	6.01	3.02	2.68
T22	n. o.	-	-	7.65	1.76	6.03	2.12	2.25

Table App 11. Chemical shifts of the protons in the structure of **LL4_mC(2,15)** at pH 7 and 5 ºC. T11, T12 and T13 could not be unambiguously assigned. na: not assigned. no: not observed.

RESIDUE	H1/H3^{+}	H42/H22	H41/H21	H6/H8	H5/Me	H1'	H2'	H2"
T1	11.49	-	-	7.76	1.83	6.45	na	na
C2	15.47	8.82	7.38	7.46	6.11	6.35	1.07	2.27
G3	10.28	8.60	5.93	8.32	-	5.95	2.94	2.66
T4	10.48	-	-	7.63	1.77	na	na	na
T5/T18	no	-	-	7.85	1.93	6.48	2.29	2.55
C6	-	8.57	7.41	7.92	6.13	6.30	2.44	
C7	15.51	10.72	8.20	7.43	5.95	6.28	0.79	2.12
G8	13.91	8.16	7.63	8.33	-	6.02	3.06	2.70
T9	10.54	-	-	7.64	1.69	5.94	2.43	2.52
T14	11.24	-	-	7.78	1.92	6.13	2.44	2.58
C15	15.51	8.60	6.96	7.49	6.33	6.41	1.10	2.20
G16	12.09	9.04	5.32	8.30	-	5.99	2.98	2.70
T17	10.51	-	-	7.64	1.80	na	na	na
C19	-	8.57	7.43	7.92	6.15	6.29	2.44	
C20	15.47	10.42	8.00	7.41	5.94	6.23	0.79	2.05
G21	13.66	8.00	7.64	8.28	-	5.99	3.03	2.66
T22	10.55	-	-	7.63	1.75	5.92	2.03	2.15

Table App 12. Chemical shifts of the protons in the structure of **LL4** at pH 7 and 5 ºC. T10, T11, T12 and T13 could not be unambiguously assigned. na: not assigned. no: not observed.

RESIDUE	H1/H3^{+}	H42/H22	H41/H21	H6/H8	H5/Me	H1'	H2'	H2"
T1	11.40	-	-	7.84	1.93	6.47	2.35	2.53
C2	15.48	8.62	7.28	7.44	6.11	6.32	1.06	2.26
G3	10.26	8.60	5.84	8.31	-	5.94	2.93	2.66
T4	no	-	-	7.62	1.75	6.13	na	na
T5/T17	no	-	-	7.84	1.94	6.13	na	na
C6	-	8.59	7.46	7.93	6.13	6.30	2.44	
C7	15.55	10.66	8.19	7.45	5.95	6.27	0.78	2.09
G8	13.94	8.20	7.63	8.32	-	6.03	3.06	2.69
T9	no	-	-	7.63	1.69	5.93	na	na
T10	no	-	-	7.76	na	5.54	na	na
T11	no	-	-	na	1.99	na	na	na
T12	no	-	-	7.84	1.99	na	na	na
T13	11.20	-	-	7.74	1.71	6.12	2.46	2.63
C14	15.55	8.66	6.97	7.49	6.39	6.41	1.07	2.17
G15	no	9.07	5.30	8.39	-	5.99	2.99	2.70
T16	no	-	-	7.64	1.81	6.08	na	na
C18	-	8.55	7.43	7.93	6.13	6.30	2.44	
C19	15.48	10.62	8.05	7.41	5.95	6.23	0.76	2.02
G20	13.65	7.96	7.67	8.27	-	5.98	3.04	2.66
T21	no	-	-	7.63	1.73	5.96	na	na

Table App 13. Chemical shifts of the protons in the structure of **LL3** at pH 7 and 5 ºC. na: not assigned. no: not observed.

RESIDUE	H1/H3^{+}	H42/H22/H10	H41/H21	H6/H8/H4	H5/Me	H1'	H2'	H2"	H3'	H4'	H5'/H5"
C1	-	8.75	7.21	7.74	5.95	6.14	2.42	2.69	4.83	n.a.	n.a.
tC^{o}2	15.02	9.82	-	7.09	-	5.91	1.11	1.73	4.68	4.39	n.a.
G3	13.93	8.80	7.03	7.96	-	5.94	2.94	2.65	4.91	4.47	3.79/3.98
T4	n.o.	-	-	7.53	1.66	6.09	1.96	2.22	n.a.	n.a.	n.a.
T5	n.o.	-	-	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
C6	-	8.75	7.67	7.89	6.12	6.26	2.51		4.81	n.a.	n.a.
C7	15.57	8.66	8.09	7.54	5.63	6.05	1.15	1.89	n.a.	4.42	n.a.
G8	13.41	n.o.	n.o.	7.98	-	5.88	2.90	2.59	4.93	n.a.	3.87/3.98
T9	n.o.	-	-	7.50	1.63	5.90	1.88	2.19	n.a.	n.a.	n.a.
T10	n.o.	-	-	7.50	1.71	6.08	2.23	2.44	n.a.	n.a.	n.a.
T11	n.o.	-	-	7.63	1.82	6.23	2.26	2.51	n.a.	n.a.	n.a.
T12	n.o.	-	-	7.67	1.88	6.31	2.30	2.40	n.a.	n.a.	n.a.
T13	n.o.	-	-	7.61	1.83	6.30	2.30	2.45	n.a.	n.a.	n.a.
C14	-	8.61	6.37	7.61	5.41	6.23	2.14	2.56	4.82	n.a.	n.a.
C15	15.57	10.83	n.o.	7.34	5.07	5.87	0.76	2.31	4.29	n.a.	n.a.
G16	n.o.	n.o.	n.o.	8.07	-	5.99	2.84	2.53	4.92	4.41	3.93/3.98
T17	n.o.	-	-	7.40	1.61	5.89	1.87	2.29	n.a.	n.a.	n.a.
T18	n.o.	-	-	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
C19	-	8.75	7.22	7.70	5.91	6.33	2.20	2.49	4.77	n.a.	n.a.
C20	15.02	11.34	8,86	7.40	5.88	6.09	0.88	2.43	4.20	n.a.	n.a.
G21	n.o.	8.23	7.21	8.13	-	6.01	2.81	2.53	4.91	4.33	3.95
T22	n.o.	-	-	7.44	1.63	5.65	1.80	2.07	4.24		n.a.

	H6	H7	H8	H9
tC^{o}2	5.98	6.21	6.67	5.92

Table App 14. Chemical shifts of **NN4_tCo(2)** at pH 7 and 5 ºC.For residues T10-T13 a tentative assignment (in grey), based on some contacts with C1 and T9, are given. na: not assigned. no: not observed.

RESIDUE	H1/H3^{+}	H42/H22/H10	H41/H21	H6/H8/H4	H5/Me	H1'	H2'	H2"	H3'	H4'
C1	-	n.o.	n.o.	7.86	6.00	6.22	2.32	2.52	4.83	4.25
C2	15.38	9.62	7.90	7.44	6.00	6.21	0.96	2.15	4.66	4.33
G3	12.13	8.22	6.76	8.25	-	5.99	2.98	2.68	5.08	
T4	n.o.	-	-	7.63	1.71	5.87	2.05	2.23	4.87	4.21
T5	n.o.	-	-	7.79	1.85	6.38	2.34	2.63	4.72	4.67
tC^{o}6	-	11.16	-	7.58	-	6.32	2.38	2.50	4.86	4.39
C7	15.40	9.79	7.73	7.39	5.87	6.21	0.88	2.13	4.65	4.34
G8	12.86	7.89	7.49	8.26	-	6.02	3.00	2.68	5.11	
T9	n.o.	-	-	7.62	1.73	5.97	2.02	2.34	4.78	
T10	n.o.	-	-	7.65	1.82	6.12	2.39	2.52	4.87	4.34
T11	n.o.	-	-	7.71	1.90	6.32	2.40	2.55	4.90	4.39
T12	n.o.	-	-	7.75	1.93	6.35	2.39	2.59	4.95	4.40
T13	n.o.	-	-	7.74	1.91	6.34	2.39	2.57	4.91	4.40
C14	-	n.o.	n.o.	7.92	6.05	6.34	2.42	2.59	4.96	4.42
C15	15.40	9.53	8.20	7.53	6.18	6.25	1.02	2.03	4.73	
G16	12.46	n.o.	n.o.	8.22	-	6.02	2.99	2.68	5.08	
T17	n.o.	-	-	7.62	1.77	6.11	2.03	2.34	4.80	
T18	n.o.	-	-	7.80	1.93	6.41	2.28	2.53	4.65	4.44
C19	-	8.46	7.54	7.70	6.07	6.26	2.31	2.43	4.78	
C20	15.38	9.69	8.12	7.47	6.00	6.25	0.93	2.08	4.72	
G21	12.92	7.92	7.49	8.27	-	6.03	3.02	2.69	8.08	
T22	n.o.	-	-	7.65	1.75	5.99	2.02	2.17	4.43	

	H6	H7	H8	H9
tC^{o}6	6.86	7.02	7.03	6.95

Table App 15. Chemical shifts of **NN4_tC^{o}(6)** at pH 7 and 5 ºC. T10-T13 a tentative assignment is given (in grey). na: not assigned. no: not observed.