Studies on Modified C- and G-rich Oligonucleotides - Towards Investigating the Role of i-Motif Structures in Telomere Biology

Hala Abou Assi

Department of Chemistry, McGill University

Montreal, Quebec, Canada

October 2017

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of ***Doctor of Philosophy***

Dedicated to my family: Mom, Dad, Fadi, Hiba, James, Hussam, and Rana. You have always been there for me through the thick and thin. I am very blessed and grateful to have such a supportive, encouraging, and caring family.

"To get the full value of joy you must
have someone to divide it with."
–Mark Twain

Copyright Statements

Most of the material in this thesis has been adapted from published and/or submitted papers and is under copyright:

Content of **Chapter 2** and **Appendix II** has been reproduced from: "Abou Assi, H.; Harkness, R.W.; Martın-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 Research*, **2016**, *44*, 4998-5009 by permission of Oxford University Press. © The Authors, 2016".

Content of **Chapter 3** has been reproduced from: "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. *Chemistry – A European Journal*, **2017**, doi: 10.1002/chem.201704591."

Content of **Chapter 4** has been reproduced from: "Abou Assi, H.; El-Khoury, R.; González, C.; Damha, M.J. 2'-Fluoroarabinonucleic acid modification traps G-quadruplex and i-motif structures in human telomeric DNA. *Nucleic Acids Research*, **2017**, *45*, 11535-11546 by permission of Oxford University Press. © The Authors, 2017".

Content of **Appendix I** has been reproduced with permission from: "Abou Assi, H.; Martinez-Montero, S.; Dixit, D.; Chua, Z.; Bohle, D.S.; Damha, M.J. Synthesis, Structure, and Conformational Analysis of Nucleoside Analogues Comprising Six-Membered 1,3-Oxathiane Sugar Rings. *European Journal of Organic Chemistry*, **2015**, 1945-1953. Copyright © 2015 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim."

Abstract

This thesis focuses primarily on studying DNA sequences with tandem G-rich and C-rich repeats that fold into G-quadruplex (G4) and i-motif structures, respectively. While G4 structures likely play a role in transcriptional regulation and telomere maintenance, much less is known about the role of i-motif structures in biological processes, especially in telomere biology. Our ability to chemically synthesize and stabilize both of these structures provides an opportunity to correlate structural and functional information of transient arrangements of G4 and i-motifs under physiological conditions. In this work, we show that replacing 2'-deoxycytidine (dC) with 2'-deoxy-2'-fluoro-arabinocytidine (2'F-araC) leads to significant stabilization of i-motifs over a wide pH range, including pH 7.0. This finding is significant because, with a few exceptions, i-motif structures unfold at pH greater than 6, making it difficult to evaluate these structures under physiological conditions. NMR experiments revealed that the stabilization provided by the 2'F-araC residues likely results from favorable electrostatic interactions that occur in the major groove of the modified C-quadruplex.

Next, we sought to determine the impact of combining the two most stabilizing sugar (2'F-araC) and nucleobase (5-methylcytidine, 5-Me-dC) modifications on the structure and stability of i-motifs. In this context, we synthesized a novel nucleoside analogue, namely 2'-deoxy-2'-fluoro-5-methyl-arabinocytidine (5-Me-2'F-araC), and incorporated it into the C-rich telomeric sequence. We found that 5-Me-2'F-araC was very well tolerated and led to structural stabilization at physiological pH similar to 2'F-araC. We also investigated the effect of combining several cytidine modifications within the same i-motif structure and determined that 5-Me-dC/2'F-araC base pairing leads to the highest thermal stability at physiological pH.

These findings paved the way to investigations concerning the possible role of telomeric i-motif structures. First, we prepared 2'F-araG modified G4 structures, which like the modified i-motifs, exhibited greater thermal stability relative to the unmodified structures. Furthermore, they were readily elongated by the enzyme telomerase in direct telomerase activity assays. Additionally, 2'F-araG and 2'F-araC substitutions allowed for the simultaneous formation of stable G4 and i-motif structures, respectively, at neutral pH. Remarkably, despite their close proximity, they remained as “trapped” G4 and i-motif structures for several weeks without

associating into a duplex. Preliminary *in vitro* assays suggested that 2'F-araC modified sequences can inhibit telomerase activity. Thus, our work shines a new light on the telomeric i-motif structure, and possibly function, and provides probes for use in the discovery of small molecule ligands and proteins that bind these structures under physiological conditions.

Résumé

Cette thèse s'articule principalement autour de l'étude de séquences d'ADN contenant des répétitions de bases guanine et cytosine en tandem conduisant respectivement à la formation de structures en G-quadruplex (G4) et en motif-i. Les structures en G4 sont probablement impliquées dans la régulation de la transcription du message génétique et dans la maintenance des télomères. En revanche, le rôle biologique des structures en motif-i est beaucoup moins clair, et ce en particulier dans la biologie des télomères. La synthèse chimique ainsi que notre capacité à stabiliser à la fois G4 et motif-i offre la possibilité d'établir un lien entre structure et fonction au sein de ces formations transitoires en milieu physiologique. Dans cet ouvrage, nous montrons que le remplacement d'un nucléoside 2'-déoxycytidine (dC) par un analogue 2'-déoxy-2'-fluoro-arabinocytidine (2'F-araC) conduit à une forte stabilisation de motifs-i sur une large plage de pH et notamment à pH neutre. La stabilisation de motifs-i à pH neutre est une avancée importante car, en dehors de quelques cas particuliers, ceux-ci se désassemblent au-delà d'un pH 6, ce qui limite drastiquement leur étude en conditions physiologiques. Des expériences menées par RMN ont permis d'attribuer l'origine de cette stabilisation à la présence d'interactions électrostatiques favorables provoquées par les nucléosides 2'F-araC et se manifestant au sein du grand sillon des motifs-i.

Ensuite, nous nous sommes intéressés à la combinaison d'analogues de nucléoside avec sucre modifié (2'F-araC) et base modifiée (5-méthylcytidine, 5-Me-dC), tous les deux stabilisateurs de motifs-i et avons souhaité déterminer l'impact de cette combinaison sur leur structure et leur stabilité. Pour ce faire, nous avons synthétisé un nouvel analogue de nucléoside, la 2'-déoxy-2'-fluoro-5-méthyl-arabinocytidine (5-Me-2'F-araC) et l'avons incorporé dans la séquence télomérique riche en cytosines, laquelle s'est retrouvée stabilisée à pH physiologique ; un effet similaire à l'incorporation de la 2'F-araC. En outre, nous avons étudié l'effet induit par la présence de multiples combinaisons de cytidine modifiées et en avons conclu que l'appariement de la 5-Me-dC et de la 2'F-araC s'accompagne d'une forte augmentation de la stabilité thermique à pH physiologique.

Ces découvertes ont posé les fondations permettant l'étude du possible rôle biologique des structures télomériques en motif-i. Ainsi, ont été préparées des structures G4 modifiées avec

des nucléosides 2'F-araG, lesquelles se sont révélées être plus stables que les G4 non modifiés. Ces structures sont des substrats de l'enzyme télomérase et leur élongation se produit naturellement lors d'un test d'activité enzymatique. Par ailleurs, la présence de substitutions 2'F-araG et 2'F-araC respectivement dans des G4 et motifs-i permet de former, simultanément, des G4 et motifs-i stables à pH neutre. Il est important de noter que ces structures se maintiennent pendant plusieurs semaines sans se réassocier en duplex, et ce malgré leur proximité. Enfin, des expériences menées *in vitro* sur des séquences modifiées avec 2'F-araC pointent vers leur action inhibitrice de la télomérase. En conclusion, le travail de cette thèse porte un regard nouveau sur la structure et peut-être aussi la fonction de la séquence télomérique en motif-i. Ce travail a, de plus, permis de concevoir des sondes utiles en milieu physiologique pour la découverte de petites molécules ou de protéines agissant comme ligands de ces structures.

Acknowledgements

First and foremost I would like to thank my mentor Dr. Masad J. Damha for being very understanding, supportive, and knowledgeable during my time in his lab. I cannot thank him enough for all the opportunities he provided me, from mentoring undergraduate students to allowing me to study abroad and develop collaborative relationships with other researchers. Despite his busy schedule as the Department head, he was always available when needed. Dr. Damha's supervision helped me become independent, able to handle several projects, plan ahead, and never give up when an experiment failed or when a paper was rejected. Again I would like to sincerely thank Dr. Damha for all his help, support, true advice and, most importantly, his genuine concern about my future and my family.

I am forever grateful for my experience at the American University of Beirut (AUB). I have many fond memories of my time at AUB. I am so thankful to have maintained close friendships over the years with many of my classmates as well as professors. I would particularly like to thank Dr. Antoine Ghauch (my undergraduate supervisor) and my dearest friends Ola, Lory, Shereen, Dina, and Sara.

I would like to thank my committee members during my yearly research reviews, Dr. Hanadi Sleiman and Dr. Karine Auclair, for their helpful feedback and constructive criticism that helped me improve and move my projects forward.

I would like to extend a special thank you to the amazing group of collaborators I was honored to work with and learn from. Dr. Carlos González, for hosting me in his lab for two months. I learned a lot from our discussions and felt I was a lab member and not just a visiting student. I am very thankful for all your help when I was in Madrid and for continuously supporting me to this day. My stay in Madrid would not have been productive and enjoyable without amazing lab members: Irene, Miguel, and Nerea. Muchas Gracias! I would also like to thank Robert Harkness (señor Mucho Bettero) and Dr. Anthony Mittermaier for their help with DSC experiments. Rob, thank you for all the fruitful G-quadruplex and thermodynamics discussions. Our science chats were a win-win situation for sure! Many thanks to Aaron Moye and Dr. Tracy Bryan at the Children's Medical Research Institute in Sydney for performing the telomerase activity assays for the G4/i-motif-duplex interconversion studies.

A special thanks to Sleiman lab members in general and Nicole, in particular, for giving me all the time I needed on their CD instrument and to Dr. Imad Baaklini in Dr. Young's lab at the Biochemistry department at McGill.

Thank you to the very helpful staff in the McGill Chemistry department: Dr. Alex Wahba and Mr. Nadim Saade for their mass spectrometry expertise, Dr. Frederick Morin and Dr. Robin Stein for their help with NMR acquisitions, Linda for the nice chats and encouraging smile before every meeting I had with Dr. Damha, Karen and Jennifer for their help with expense reports, and last but not least, Chantal for her continuous help from before joining the program till the very last day.

Present and past Damha lab members: Glen, Matt, Maryam Y, Maryam H, Dilip, Elise, Jovanka, Logan, James, Sunit (the king of crystals) for always having positive and encouraging feedback, Adam for your caring advice and for literally being an encyclopedia, and Dan for the bootcamps and Saturday brunches in the lab. Danielle, we've been through this journey together from day one. Your thoughtfulness and creativity made every occasion (birthday, defense, farewell…) a memorable one. I can never forget your super powers in organizing, your eagle eye in editing a slide or a paper, and the delicious sweets you prepared for the lab. Saul, my Spanish friend, thank you for not only being a very helpful and knowledgeable postdoc, but also for being a dear friend who has always encouraged me throughout my Ph.D. Ken, you are one of the most helpful scientists I've ever met, I definitely learned a lot from you. Jory thank you for being very helpful, great at planning activities, having a great sense of humor, and for translating the abstract of my thesis. Leonora (my "deer" daughter) you were a great addition to our lab and a perfect fit. Your presence and sense of humor made everyone around you happy, your smile is contagious. Roberto (the coffee boy) I am very glad you joined our lab; I can finally speak Arabic in the lab and say Arabic proverbs without the need to translate (although you didn't understand most of them)! I am sure my project will be in good hands! Jimmy, I had the pleasure to work with you during your honors project in our lab. We had a great and productive time working together!

My friends from inside and outside the department that made Montreal feel like home away from home. Amani Hariri you are the reason for me joining McGill. I can't forget your

words “throw an application and then think about it” and here I am today, six years later, writing my Ph.D. thesis. Carolin Madwar you were the best morning partner I could ask for; although going to gym with you was super tough, we enjoyed every single moment and it was the best way to start the day strong. Amani and Carolin, your friendship is one of the best experiences at McGill. Nidal and Hoda thank you for being there for me during my first months in Montreal and for helping me around. A special thank you to the dearest friends who became my family in Montreal: the El-Oud family (Mira, Bassam, and Lanloon) and the Shaars (Nagham, Wissam, and Yasmeena). Seeing you every weekend was a must; your presence gave life in Montreal a different flavor. I can’t forget to thank Antisar for her care, Keena-Lou for the tough workouts and positive vibes, Serene Bayram for the sweet smile, and Diana Madwar, Ghada Ayoub, Hassan Fakih, and Hiba Momneh for their friendships.

I am very grateful for all the support I had from my family. I owe every success to you. It is through your extraordinary support that I was able to get to this level in life. My mom (Nabila) and dad (Najib), you have sacrificed all your life to make sure we received the best education in the top schools and universities. I know you have been waiting for my defense day for years and it has been your dream, which I worked hard in making it come true. My sister (Hiba) and brother-in-law (James) you were my number one supporters when I first came to Montreal. Although I used to get very annoyed when people asked me “are you as smart as your sister”, now I would like to thank you for being my role model and the hard-working, smart lady I always looked up to. My brother (Hussam) and his family (Rana, Rima and Ziad) who always brought a smile in every facetime call. Fadi, my husband, whom I met in Montreal while doing my Ph.D. and who was always by my side since day one, through the ups and downs, regardless of the distance. Thanks for all the rehearsals I did on skype and for all the nights you stayed up late to call me on my way home after a long day in the lab. I am forever grateful for all the patience you had (and continue to have!). Most importantly, you always believed in my potential, always encouraged me, and helped boost my self-esteem.

“We all take different paths in life, but no matter where we go, we take a little of each other everywhere.”
–Tim McGraw

Preface

Prof. Masad J. Damha provided the funding and the intellectual guidance for all the projects presented in this thesis.

In **Chapter 2**, Robert W. Harkness carried out the DSC experiments under the supervision of Dr. Anthony K. Mittermaier. Dr. Nerea Martin-Pintado performed the preliminary 1D NMR experiments on the hexanucleotide sequences in the laboratory of Dr. Carlos González (Madrid, Spain). Dr. González hosted me in his lab for two months to run the 1D and 2D NMR experiments under his supervision.

In **Chapter 3**, Yu Chen Lin (Jimmy) assisted in synthesizing 5-Me-2'F-araC during his undergraduate honors project. Israel Serrano from Dr. González's lab solved the NMR structure of HJ-2.

In **Chapter 4**, Roberto El-Khoury assisted in sample preparation for 2D NMR analysis and all the NMR experiments were carried out in Dr. González's lab. Aaron Moye from Dr. Tracy Bryan's lab at the Children's Medical Research Institute in Sydney conducted the telomerase activity assays.

In **Appendix I**, Dr. Saul Martinez-Montero helped with NMR interpretations, Dr. Dilip Dixit helped with synthesis designs and Dr. Zhijie Chua from Dr. Scott Bohle's lab solved the crystal structure of the 1,3-oxathiane nucleoside analogue.

Table of Contents

Copyright Statements ..... **ii**

Abstract ..... **iii**

Résumé ..... **v**

Acknowledgements ..... **vii**

Preface ..... **x**

Table of Contents ..... **xi**

List of Figures ..... **xvii**

List of Tables ..... **xxiii**

List of Schemes ..... **xxiv**

List of Abbreviations ..... **xxv**

Chapter 1 Introduction **1**

1.1 Nucleic Acids: The Building Blocks of Life and Beyond 2

1.2 Nucleic Acid Structure and Function 2

1.2.1 Natural Nucleosides and Nucleotides 3

1.2.2 Conformation of Nucleosides and Nucleotides 4

1.2.3 The DNA Double Helix 6

1.2.4 The *Central Dogma*: The Mechanism of Gene Expression 8

1.3 Chemically-Modified Nucleosides and Oligonucleotides 9

1.3.1 Modified Nucleosides as Therapeutic Agents 9

1.3.2 Oligonucleotides as Gene Silencing Therapeutics 12

1.3.3 2'-Deoxy-2'-Fluoro-β-D-Arabinonucleic Acid (2'F-ANA) 15

1.4 Beyond the DNA Double Helix and Watson-Crick Base-Pairing 16

1.5 Human Telomere Biology: Telomeres and Telomerase in Cellular Aging and Cancer 18

1.5.1 Telomeres: The Protective Caps of Chromosomes 18

1.5.2 Telomerase and Telomere Maintenance 20

1.6 G-Quadruplexes: Structure, Function, and Biological Relevance 22

1.6.1 G-tetrads: The Building Blocks of G4 Structures 22

1.6.2 G-Quadruplexes: A Single Building Block Leading to a High Degree of Structural Polymorphism 23

1.6.3 Trapping G-Quadruplex Geometries Using Chemically-Modified Nucleotides .. 25
1.6.4 Location of G-Quadruplex-Forming Sequences Throughout the Genome 28
1.6.5 Biological Significance of G-Quadruplex Structures 29
1.6.6 Ligands and Proteins that Recognize and Bind to G-Quadruplex Structures 31
1.7 Nucleic Acid i-Motifs: Structure, Stability, and Targeting Ligands 32
1.7.1 i-Motif Structure and Topology 33
1.7.2 Factors Affecting the Stability of i-Motif Structures 35
1.7.2.1 C-tract Length and Genomic Analysis 36
1.7.2.2 Effects of Chemical Modifications on i-Motif Stability 38
1.7.2.2.1 Sugar Modifications 38
1.7.2.2.2 Nucleobase Modifications 40
1.7.2.2.3 Backbone Modifications 41
1.7.2.3 The Nature and Length of Connecting Loops 42
1.7.2.4 Impact of Ionic Strength, Molecular Crowding, and Negative Supercoiling 42
1.7.3 Interaction with Ligands and Proteins 43
1.7.3.1 Ligands 43
1.7.3.2 Proteins 45
1.7.4 Biological Relevance of i-Motifs *in vivo* – Is There One? 47
1.7.4.1 Inhibition of Telomerase Activity 47
1.7.4.2 Transcriptional Regulation of Gene Expression 48
1.7.4.3 Regulation of DNA Biosynthesis 49
1.7.4.4 Mutual Exclusivity of i-Motifs and G-Quadruplexes and its Effect on Gene Expression 50
1.8 Thesis Objectives 54
1.9 References 56
Chapter 2 Stabilization of i-Motif Structures by 2'-β-Fluorination of DNA 78
2.1 Introduction 79
2.2 The Effect of 2'F-araC Substitutions on Thermal and pH Stability of i-Motif Structures 81
2.3 Studying i-Motif Formation by ^{1}H-NMR 87

2.4 Determining the Molecularity of the i-Motif Structures *via* Gel Electrophoretic Experiments and ^{19}F-NMR 91
2.5 Thermodynamic Analysis by Differential Scanning Calorimetry 95
2.6 Structural Basis of 2'F-araC Stabilization 99
2.7 Significance of Results 104
2.8 Conclusions 106
2.9 Experimental Section 106
2.9.1 Oligonucleotide Synthesis and Purification 106
2.9.2 UV melting Studies 107
2.9.3 Circular Dichroism 108
2.9.4 Gel Electrophoresis 108
2.9.5 NMR Experiments and Constraints 108
2.9.6 Structural Determination 109
2.9.7 Differential Scanning Calorimetry 109
2.10 References 110
Chapter 3 Probing Synergistic Effects of DNA Methylation and 2'-β-Fluorination on i-Motif Stability 115
3.1 Introduction 116
3.2 Synthesis of 2'-Deoxy-2'-Fluoro-5-Methyl-Arabinocytidine Phosphoramidite 117
3.3 Studying the Effect of Nucleobase and Sugar Modifications on i-Motif Formation 120
3.4 Exploring the Effect of Chemical Modifications on the pK_a of Nucleosides and $pH_{1/2}$ of i-Motif Structures 123
3.5 Structural Determination *via* 2D NMR Spectroscopy 126
3.6 Conclusion 132
3.7 Experimental Data 132
3.7.1 General Considerations 132
3.7.2 Experimental Procedure for All Synthesized Compounds 133
3.7.2.1 2-deoxy-2-fluoro-3,5-di-O-benzoyl-α-D-arabinofuranosyl bromide (2) 133
3.7.2.2 N^4-benzoyl-5-methylcytosine (4) 133
3.7.2.3 N^4-benzoyl-5-methyl-1-[2-deoxy-2-fluoro-3,5-di-*O*-benzoyl-β-arabinofuranosyl]cytosine (5) 134

3.7.2.4 5-methyl-1-[2-deoxy-2-fluoro-β-D-arabinofuranosyl]cytosine (6) 134
3.7.2.5 1-[2-deoxy-2-fluoro-3,5-di-*O*-benzoyl-β-D-arabinofuranosyl] thymine (8) 135
3.7.2.6 5-methyl-1-[2-deoxy-2-fluoro-β-D-arabinofuranosyl]cytosine (6) 136
3.7.2.7 N^4-benzoyl-5-methyl-1-[2-deoxy-2-fluoro-β-D-arabinofuranosyl] cytosine (10) 137
3.7.2.8 N^4-benzoyl-5-methyl-1-[2-deoxy-2-fluoro-5-(4,4'-dimethoxytrityl)-β-D-arabinofuranosyl]cytosine (11) 137
3.7.2.9 N^4-benzoyl-5-methyl-1-[3-(2-cyanoethoxy(diisopropylamino) -phosphinyl)-2-deoxy-2-fluoro-5-(4,4'- dimethoxytrityl)-β-D arabinofuranosyl]cytosine (12) 138
3.7.3 Oligonucleotide Synthesis and Purification 138
3.7.4 UV Melting Studies 139
3.7.5 Circular Dichroism Studies 140
3.7.6 Nuclear Magnetic Resonance 140
3.7.7 Structural Determination 140
3.8 References 141
Chapter 4 2'-Fluoroarabinonucleic Acid Modification Traps G-Quadruplex and i-Motif Structures in Human Telomeric DNA 145
4.1 Introduction 146
4.2 Biophysical Properties of 22mer Cytosine and Guanine-Rich Single Strands and Duplexes 149
4.2.1 Structural Properties of 22mer Cytosine and Guanine-Rich Telomeric Single Strands 149
4.2.2 Structural Properties of Cytosine and Guanine-Rich Telomeric Duplexes (Annealing *versus* Mixing) 152
4.2.3 NMR Kinetic Experiments of 22mer Mixes 157
4.3 Telomerase Activity Assays 160
4.4 Properties of 35mer Cytosine and Guanine-Rich Single Strands and Duplexes 164
4.4.1 NMR Kinetic Experiments of 35mer Mixes 167
4.5 Conclusions 172

4.6 Experimental Data 173
4.6.1 Oligonucleotides Synthesis and Purification 173
4.6.2 Thermal Melting Experiments 174
4.6.3 Circular Dichroism 175
4.6.4 NMR Kinetic Experiments 175
4.6.5 Gel Electrophoresis Experiments 175
4.6.6 Telomerase Activity Assays 176
4.7 References 177
Chapter 5 Contributions to Knowledge and Future Work 182
5.1 Summary and Contributions to Knowledge 183
5.1.1 Stabilization of i-Motif Structures by 2'-β-Fluorination of DNA (Chapter 2) 183
5.1.2 Probing Synergistic Effects of DNA Methylation and 2'-β-Fluorination on i-Motif Stability (Chapter 3) 183
5.1.3 2'-Fluoroarabinonucleic Acid Modification Traps G-Quadruplex and i-Motif Structures in Human Telomeric DNA (Chapter 4) 184
5.1.4 Synthesis, Structure, and Conformational Analysis of Nucleoside Analogues Comprising Six-Membered 1,3-Oxathiane Sugar Rings (Appendix I) 184
5.2 Suggestions for Future Work 185
5.3 List of Publications 188
5.4 List of Conference Presentations 189
Appendix I Synthesis, Structure, and Conformational Analysis of Nucleoside Analogues Comprising Six-Membered 1,3-Oxathiane Sugar Rings 190
AI.1 Introduction 191
AI.2 Synthesis of 1,3-Oxathiane Nucleoside Analogues 193
AI.3 Crystal Structure and Conformational Analysis of 1,3-Oxathiane Nucleoside Analogues 199
AI.4 Conclusions 201
AI.5 Experimental Section 201
AI.5.1 2-benzoyloxymethyl-1,3-oxathiane (6) 202
AI.5.2 2-benzoyloxymethyl-1,3-oxathian-3-one (7) 202
AI.5.3 2-benzoyloxymethyl-4-acetoxy-1,3-oxathiane (8) 203

AI.5.4 General Procedure for the Glycosylation of Compound 8 with Pyrimidine Bases Followed by Deprotection to Afford Compounds 10a-c.....203
AI.5.4.1 1-(2-benzoyloxymethyl-1,3-oxathian-4-yl)-cytosine (9a).....204
AI.5.4.2 1-(2-hydroxymethyl-1,3-oxathian-4-yl)-cytosine (10a).....204
AI.5.4.3 1-(2-benzoyloxymethyl-1,3-oxathian-4-yl)-thymine (9b).....205
AI.5.4.4 1-(2-hydroxymethyl-1,3-oxathian-4-yl)-thymine (10b).....205
AI.5.4.5 1-(2-benzoyloxymethyl-1,3-oxathian-4-yl)-5-fluorocytosine (9c).....206
AI.5.4.6 1-(2-hydroxymethyl-1,3-oxathian-4-yl)-5-fluorocytosine (10c).....206
AI.5.5 General procedure for the synthesis of purine 1,3-oxathiane nucleosides.....207
AI.5.5.1 9-(2-benzoyloxymethyl-1,3-oxathian-4-yl)-6-chloropurine (11).....207
AI.5.5.2 7-(2-benzoyloxymethyl-1,3-oxathian-4-yl)-6-chloropurine (12).....208
AI.5.5.3 9-(2-hydroxymethyl-1,3-oxathian-4-yl)-adenine (13).....208
AI.5.5.4 7-(2-hydroxymethyl-1,3-oxathian-4-yl)-adenine (14).....209
AI.5.5.5 α-7-(2-hydroxymethyl-1,3-oxathian-4-yl)-6-methoxypurine (15).....209
AI.5.5.6 β-7-(2-hydroxymethyl-1,3-oxathian-4-yl)-6-methoxypurine (16).....210
AI.5.6 X-ray Crystallography.....210
AI.6 References.....211
Appendix II: Supplementary Information for Chapter 2.....215
AII.1 Differential Scanning Calorimetry Fitting Procedure.....215
AII.2 $T_{1/2}$ Data 218
AII.3 CD Spectra.....219
AII.4 ^{1}H-NMR Spectra for Hexamer Sequences.....220
AII.5 ^{19}F-NMR Spectra for Hexamer Sequences.....222
AII.6 DSC Data.....225
AII.7 NMR Data.....226
AII.8 References.....230

List of Figures

Figure 1.1: Nucleic acids components.....4
Figure 1.2: The pseudorotational wheel and sugar pucker notation of the furanose ring.....5
Figure 1.3: B-form dsDNA and A-form dsRNA.....7
Figure 1.4: Activation of nucleoside analogues into their 5'-triphosphorylated form.....9
Figure 1.5: Structures of some FDA approved nucleoside analogues.....10
Figure 1.6: Chemical structure of sofosbuvir and tenofovir alafenamide (TAF).....12
Figure 1.7: Chemical structures of some modifications utilized in ON therapeutics.....14
Figure 1.8: Chemical structure of 2'F-ANA, 2'F-RNA, Clofarabine, and 2'β-fluoro-tricyclo nucleic acids (2'F-tc-ANA).....16
Figure 1.9: Some non-canonical hydrogen-bonding interactions.....17
Figure 1.10: Representation of the telomeric region and the association of telomerase and shelterin component proteins to telomere ends.....19
Figure 1.11: Mechanism of telomere elongation by telomerase.....21
Figure 1.12: G-quadruplex parameters and topologies.....24
Figure 1.13: A) The guanine in 2'F-araG adopts an anti glycosidic conformation. B) Distances in the crystallographic structure between H2'−H8 and H2"−O4'. C) Distances between F−H8 and H2"−O4' in the 2'F-araG modified structure. D) Scheme of the parallel propeller G4 with all guanines in the anti conformation.....27
Figure 1.14: Immunofluorescence for BG4 antibodies on metaphase chromosomes isolated from Hela cervical cancer cells leading to the observation of G4 structures in chromosomes.....29
Figure 1.15: A) Inhibition of telomerase activity due to G4 formation. B) Proposed mechanism for the elongation of parallel intermolecular G4 structures by telomerase.....31
Figure 1.16: A) $C{\cdot}C^+$ base pairs. B) 3D structure of a tetramolecular i-motif. C) Stacking between intercalating $C{\cdot}C^+$ base pairs. D) 3'E intercalation topology. E) 5'E intercalation topology.....34
Figure 1.17: Sugar modifications introduced in i-motif structures.....38
Figure 1.18: The orientation of ribose residues in the narrow groove of an i-motif.....39

Figure 1.19: Structure of i-motif binding ligands.....44
Figure 1.20: A) The conformational equilibrium between the hairpin and i-motif structures of the C-rich strand in the BCL-2 promoter. B) IMC-76 binds to the hairpin structure leading to transcriptional repression. C) IMC-48 binds to the central loop of the BCL-2 i-motif providing the binding sites for RRMs 1 and 2 of the hnRNP LL. D) hnRNP LL binds to the unfolded C-rich strand causing transcriptional activation of BCL-2.....45
Figure 1.21: The effect of carboxylated SWNTs on telomerase inhibition and telomere uncapping.....48
Figure 1.22: The proposed mechanism of unwinding of structured DNAs by the Klenow fragment of DNA polymerase.....50
Figure 1.23: The proposed equilibrating forms of the NHE III_1 of the c-MYC promoter produced under negative supercoiling.....52
Figure 1.24: The proposed molecular mechanosensor mechanism for the differential control of MYC expression through the NHE III_1.....54
Figure 2.1: A) $C{\cdot}C^+$ and 2'F-ara$C{\cdot}C^+$ base pairs. Schematic representation of B) the tetrameric i-motif formed from dTCCCCC (H-1), C) the human telomeric repeat (HT-0), and D) the centromeric A-box (HC-0).....80
Figure 2.2: UV-melting curves for centromeric and telomeric sequences at pH 5.0 and 7.0.....83
Figure 2.3: A) and B) pH-dependent CD spectra for H-1 and H-4, respectively. C) and D) CD spectra for centromeric and telomeric sequences at pH 7.0, respectively.....85
Figure 2.4: Nucleoside pK_a and i-motif $pH_{1/2}$ data.....86
Figure 2.5: ^{1}H-NMR melting experiments for H-1 and H-4 at pH 6.0 and pH 7.0.....88
Figure 2.6: ^{1}H-NMR melting experiments for centromeric sequences at pH 7.0.....89
Figure 2.7: ^{1}H-NMR melting experiments for telomeric sequences at pH 7.0.....90
Figure 2.8: Native PAGE experiments for H-1, H-2, H-3, H-4, H-5, and H-6.....91
Figure 2.9: Concentration-dependent native gels for the tetramolecular i-motifs.....92
Figure 2.10: Denaturation profiles of H-1 and H-6.....93
Figure 2.11: ^{19}F-NMR melting experiments for H-6 5'-Tf(CCCCC)-3', pH 5.0.....94

Figure 2.12: ^{19}F-NMR spectra for H-6 5'-Tf(CCCCC)-3' recorded after slow and fast annealing procedures.....95
Figure 2.13: ^{1}H-NMR spectra for H-6 at pH 5.0 (diluted versus concentrated) 95
Figure 2.14: Excess C_p profiles for the control and 2'F-araC modified tetrameric i-motif structures.....97
Figure 2.15: $\Delta T_{1/2}$ as a function of the number of 2'F-araC modifications.....98
Figure 2.16: Amino-H2'/2", H1'-aromatic, and H1'-H1' regions of the NOESY spectra of H-4 at pH 5.0.....100
Figure 2.17: DQF-COSY superimposed on the NOESY spectra of H-4 at pH 5.0.....101
Figure 2.18: Average solution structure and schematic representation of the major 3'E species of H-4 at pH 5.0 and pH 7.0.....102
Figure 2.19: Imino-amino region of the NOESY spectra of H-4 at different pHs.....102
Figure 2.20: NOESY spectra of H-4 at neural pH.....103
Figure 2.21: Details of sugar contacts in 2'F-araC substituted i-motif (H-4) and in an unmodified i-motif (H-1).....105
Figure 3.1: Base and sugar modifications previously introduced in i-motif structures.....117
Figure 3.2: NOESY NMR spectrum of 5-methyl-1-[2-deoxy-2-fluoro-β-D-arabinofuranosyl]cytosine (6).....119
Figure 3.3: ^{31}P-NMR spectrum of amidite 12 showing the long-range ^{19}F-^{31}P coupling.....119
Figure 3.4: Thermal melting curves for the modified structures at pH 5.0, 7.0, and 7.4.....121
Figure 3.5: CD spectra recorded at pH 5.0 and pH 7.0 for the 5-Me-dC modified sequences.....122
Figure 3.6: UV spectra for the free nucleosides as a function of pH.....124
Figure 3.7: CD spectra for HT-0 and HJ-2 reported from pH 4.0 to 9.0.....125
Figure 3.8: Plot of ellipticity at 284 nm as a function of pH.....126
Figure 3.9: A) ^{1}H-NMR melting experiments of HJ-2 at pH 7.0. B) H1'-H1' region of the NOESY spectra of HJ-2 at pH 7.0. C) Schematic representation of HJ-2 structure with 5'E intercalation topology. D) Two views of the solution structure of HJ-2.....128
Figure 3.10: Regions of the NOESY spectra of HJ-2 at pH 7.0.....129

Figure 3.11: Regions of the TOCSY spectra of HJ-2 and schematic representation of 2'F-araC·5-Me-dC$^+$ base pair.....130
Figure 3.12: Details of sugar contacts in an unmodified telomeric i-motif (PDB 1ELN) and in 2'F-araC and 5-Me-dC substituted i-motif structure (HJ-2).....131
Figure 4.1: A) A·T and C·G Watson-Crick base pairs and schematic representation of an antiparallel duplex. B) G-quartet stabilized by monovalent cation and schematic representation of an intramolecular G-quadruplex. C) C·C$^+$ base pair and schematic representation of an intramolecular i-motif.....147
Figure 4.2: Structures of 2'F-araC, 2'F-araG, and 2'F-riboG.....147
Figure 4.3: Models representing G4/i-motif-duplex interconversion.....149
Figure 4.4: ^{1}H-NMR melting temperature experiments for 22G0, 22G1, and 22G3.....150
Figure 4.5: CD spectra for 22G-rich and 22C-rich sequences at pH 7.0.....151
Figure 4.6: UV-melting curves for the 22G-rich strands.....151
Figure 4.7: UV-melting curves for the 22mer duplexes.....153
Figure 4.8: CD spectra as a function of pH for the 22mer duplexes.....155
Figure 4.9: ^{1}H-NMR melting temperature experiments for the 22mer duplexes.....156
Figure 4.10: ^{1}H-NMR kinetic experiments for the 22mer mixes at 5 °C.....158
Figure 4.11: ^{1}H-NMR kinetic experiments for the 22mer mixes at 25 °C.....159
Figure 4.12: Telomerase activity assay at pH 7.5 and 37 °C.....161
Figure 4.13: Telomerase activity assay at pH 7.5 and 25 °C.....163
Figure 4.14: CD spectra of 35G-rich and 35C-rich sequences.....165
Figure 4.15: ^{1}H-NMR melting experiments for 35G0, 35G3, and 35C2.....165
Figure 4.16: UV-melting curves and CD spectra for the 35mer duplexes.....166
Figure 4.17: ^{1}H-NMR melting temperature experiments for the 35mer duplexes.....167
Figure 4.18: ^{1}H-NMR kinetic experiments for the 35mer mixes.....168
Figure 4.19: Schematic representation of the proposed scenarios: Scenario A) a duplex moiety co-exists in the same complex with G4 and i-motif. Scenario B) a duplex moiety coexists with a C-rich ssDNA and a G4 (B1), a G-rich ssDNA and an i-motif (B2). C) Imino region of the ^{1}H-NMR spectrum of 35G3:35C2.....169
Figure 4.20: Native PAGE experiments for the 35mer duplexes and 35mer mixes.....170

Figure 4.21: A) Two regions of the NOESY spectra of 35C2. B) Enlarged H5-H6 region of the TOCSY spectra of 35C2 and the 1:1 complex 35G3:35C2.....171
Figure 5.1: Proposed single molecule FRET experiments.....186
Figure 5.2: Proposed designs for G4 and i-motif structures on the complementary strands of telomeric and oncogene promoter regions.....188
Figure AI.1: Clinically approved nucleoside reverse transcriptase inhibitors.....191
Figure AI.2: Previously reported six-membered nucleoside analogues.....192
Figure AI.3: RP-HPLC separation of anomeric mixtures of nucleoside 10a.....194
Figure AI.4: NOESY NMR spectrum of 9-(2-hydroxymethyl-1,3-oxathian-4-yl)-adenine (13).....196
Figure AI.5: NOESY NMR spectrum of β-7-(2-hydroxymethyl-1,3-oxathian-4-yl)-6-methoxypurine (16).....197
Figure AI.6: HMBC NMR spectrum of 9-(2-hydroxymethyl-1,3-oxathian-4-yl)-adenine (13).....197
Figure AI.7: HMBC NMR spectrum of β-7-(2-hydroxymethyl-1,3-oxathian-4-yl)-6-methoxypurine (16).....198
Figure AI.8: UV spectra of nucleosides 13 and 14 in comparison with riboadenosine.....198
Figure AI.9: Crystal structure of 7-[(2S, 4S)-2-hydroxymethyl-1,3-oxathian-4-yl]-6-methoxypurine (16).....199
Figure AI.10: The two possible conformations of 1,3-oxathiane nucleosides 15 and 16 in solution.....200
Figure AI.11: Crystal structure of the two independent molecules of nucleoside 16 in asymmetric unit.....211
Figure AI.12: Crystal packing of nucleoside 16.....211
Figure AII.1: CD spectra of hexamer sequences at pH 5.0-7.5.....219
Figure AII.2-a: ^{1}H-NMR melting experiments for H-2 5'-TCC(fC)CC-3'.....220
Figure AII.2-b: ^{1}H-NMR melting experiments for H-3 5'-TC(fC)(fC)CC-3'.....220
Figure AII.2-c: ^{1}H-NMR melting experiments for H-4 5'-TCC(fC)(fC)C-3'.....221
Figure AII.2-d: ^{1}H-NMR melting experiments for H-5 5'-TC(fC)(fC)(fC)C-3'.....221
Figure AII.2-e: ^{1}H-NMR melting experiments for H-6 5'-Tf(CCCCC)-3'.....222
Figure AII.3-a: ^{19}F-NMR melting experiments for H-2 5'-TCC(fC)CC-3'.....222

Figure AII.3-b: ^{19}F-NMR melting experiments for H-3 5'-TC(fC)(fC)CC-3'.....223
Figure AII.3-c: ^{19}F-NMR melting experiments for H-4 5'-TCC(fC)(fC)C-3'.....223
Figure AII.3-d: ^{19}F-NMR melting experiments for H-5 5'-TC(fC)(fC)(fC)C-3'.....224
Figure AII.4: Raw DSC power data for unmodified and modified tetrameric i-motif structures.....225
Figure AII.5: Tetrameric structure of H-4 5'-TCC(fC)(fC)C-3' at acidic pH.....226
Figure AII.6: NMR assignment of the major species of H-4 at neural pH.....227
Figure AII.7: ^{19}F-^{1}H HOESY spectrum of H-4.....228

List of Tables

Table 1.1: Average helical parameters for B-form dsDNA and A-form dsRNA.....8
Table 1.2: Characteristic features of B-DNA duplexes, G4s, and i-motifs.....51
Table 2.1: $T_{1/2}$ values for the (un)modified sequences.....82
Table 2.2: $T_{1/2}$ values for the hexamer sequences at different pH values.....84
Table 2.3: Nucleoside pK_a and i-motif $pH_{1/2}$ values.....87
Table 2.4: Thermodynamic parameters extracted from fits of the tetrameric i-motif DSC melting experiments.....98
Table 2.5: MS characterization of the studied sequences.....107
Table 3.1: UV-melting data of the modified oligonucleotide sequences.....120
Table 3.2: Nucleoside pK_a and i-motif $pH_{1/2}$ values.....125
Table 3.3: Chemical shifts of HJ-2, pH 7.0, 10 mM sodium phosphate buffer, T = 5 °C....131
Table 3.4: MS characterization of the studied sequences.....139
Table 4.1: Thermal stability of the telomeric 22G and 22C-rich sequences (pH 7.0).....148
Table 4.2: Melting temperatures of the telomeric 22G-rich and 22C-rich sequences (pH 7.0) at different concentrations.....152
Table 4.3: Melting temperatures of 22mer human telomeric duplexes at different pH.....154
Table 4.4: Melting temperatures of the telomeric 22G-rich and 22C-rich single strands at different pH values.....154
Table 4.5: Melting temperatures of the telomeric 35G-rich and 35C-rich sequences.....164
Table 4.6: Melting temperatures of 35mer human telomeric duplexes at different pHs.....166
Table 4.7: MS characterization of the studied sequences.....174
Table 5.1: Sequences to be utilized in telomerase activity assays.....187
Table AI.1: ^{1}H-^{1}H Coupling Constant Values for Nucleoside (15) and Nucleoside (16).....200
Table AII.1: $T_{1/2}$ values for the (un)modified sequences at pH 5.0.....218
Table AII.2: Assignment list of H-4 at pH 5.0/6.0, T = 5 °C, [H-4] = 2.5 mM.....228
Table AII.3: Experimental constraints and calculation statistics of H-4.....229
Table AII.4: Average dihedral angles and order parameters of H-4.....229

List of Schemes

Scheme 3.1: Synthetic scheme representing the synthesis of 5-Me-2'F-araC phosphoramidite 12.....118
Scheme AI.1: Synthesis of pyrimidine 1,3-oxathiane nucleosides.....194
Scheme AI.2: Synthesis of 1,3-oxathiane purine nucleosides.....195

List of Abbreviations

1D	one dimensional
2D	two dimensional
2'F-ANA	2'-deoxy-2'-fluoroarabinonucleic acid
2'F-araC	2'-deoxy-2'-fluoro-arabinocytidine
2'F-araG	2'-deoxy-2'-fluoro-arabinoguanosine
2'F-riboG	2'-deoxy-2'-fluoro-guanosine
2'F-RNA	2'-deoxy-2'-fluororibonucleic acid
3TC	Lamivudine
5-hMe-dC	5-hydroxymethylcytidine
5-Me-2'F-araC	2'-deoxy-2'-fluoro-5-methyl-arabinocytidine
5-Me-dC	5-methylcytidine
A	adenosine
Å	Angstrom
A_{260}	UV absorbance at 260 nm
ABC	Abacavir
ACN	acetonitrile
ANA	arabinonucleic acid
AON	antisense oligonucleotides
AZT	Zidovudine
Bcl-2	B-cell lymphoma 2
bp	base pair
BPE	base-pairing energy
BSA	bis(trimethylsilyl)acetamide
Bz	benzoyl
BzCl	benzoyl chloride
C	cytidine
^{o}C	celsius
CD	circular dichroism
COSY	correlation spectroscopy

C_p	heat capacity
CPG	controlled pore glass
Cpm	counts per minute
CRISPR	clustered regularly interspaced short palindromic repeats
d	doublet (NMR)
d	days (reaction time)
D-loop	displacement loop
d4T	Stavudine
d4U	2',3'-didehydro-2',3'-dideoxyuridine
dC	2'-deoxycytidine
DCM	dichloromethane
dd	doublet of doublets
ddC	Zalcitabine
ddd	doublet of doublet of doublets
ddI	Didanosine
ddN	2',3'-dideoxynucleoside
ddU	2',3'-dideoxyuridine
DEPC	diethylpyrocarbonate
DIPEA	N,N-diisopropylethylamine
DMAP	4-dimethylaminopyridine
DMF	dimethylformamide
DMSO	dimethylsulfoxide
DMTr	4,4'-dimethoxytrityl
DNA	2'-deoxyribonucleic acid
dNTPs	deoxynucleoside triphosphates
DQF-COSY	double-quantum filtered COSY
DSC	differential scanning calorimetry
dsDNA	double-stranded DNA
dsRNA	double-stranded RNA
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid

ESI	electrospray ionization
Et_3N	triethylamine
EtOH	ethanol
FDA	Food and Drug Administration
FRET	Förster resonance energy transfer
FTC	Emtricitabine
G	guanosine
G4	G-Quadruplex
ΔG	Gibbs free energy
GR	G-register
h	hour (reaction time)
ΔH	enthalpy
%H	percent hyperchromicity $(A_{final}-A_{initial})/A_{initial}$
HAART	highly active antiretroviral therapy
HBV	hepatitis B virus
HC	human centromeric sequence
HCV	hepatitis C virus
HCMV	human cytomegalovirus
HIV	human immunodeficiency virus
HMBC	heteronuclear multiple bond coherence
HMDS	hexamethyldisilazane
hnRNP	heterogeneous nuclear ribonucleoprotein
HOESY	heteronuclear overhauser effect spectroscopy
HPLC	high performance liquid chromatography
HRMS	high-resolution mass spectra
HSQC	heteronuclear single quantum coherence
HSV	herpes simplex virus
HT	human telomeric sequence
hTERT	human telomerase reverse transcriptase
hTR	human telomerase RNA component
IFN	Interferon

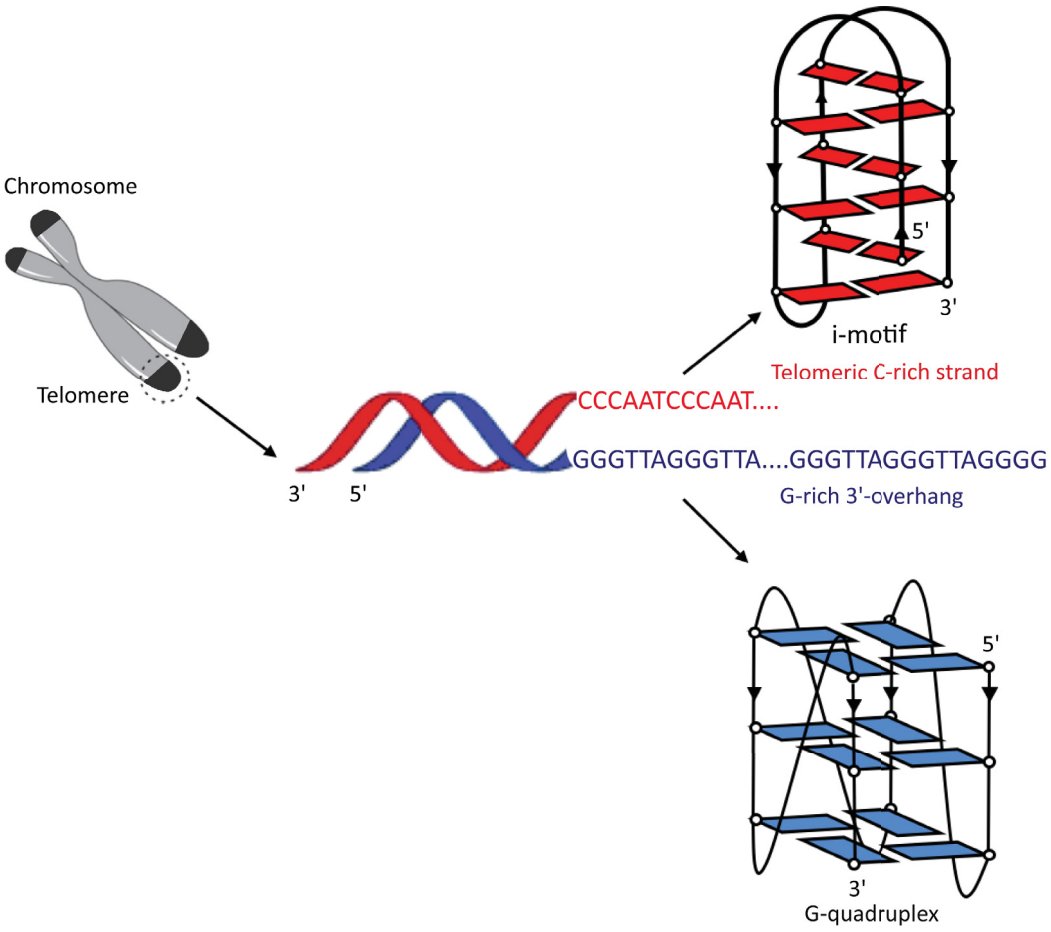
IL-6	Interleukin 6
J	coupling constant (Hz)
k	reaction rate constant
kb	kilobase
KCl	potassium chloride
KF	Klenow fragment
KP_i	potassium phosphate buffer
KRAS	Kirsten rat sarcoma
LNA	locked nucleic acid
MeOH	methanol
MD	molecular dynamics
m	multiplet
min	minutes
mM	mili-molar
µM	micro-molar
mRNA	messanger RNA
MS	mass spectrometry
NA	nucleoside analogue
NaOMe	sodium methoxide
NaP_i	sodium phosphate buffer
NHE	nuclease-hypersensitive element
nM	nano-molar
NMR	nuclear magnetic resonance
NOE	nuclear overhauser effect
NOESY	nuclear overhauser effect spectroscopy
NRTI	nucleoside reverse transcriptase inhibitor
nt	nucleotide
ON	oligonucleotide
OTE	off-target effect
P	phase angle
PAGE	polyacrylamide gel electrophoresis

PBMC	peripheral blood mononuclear cells
PCBP	poly-C-binding proteins
PDB	protein data bank
PDGFR- β	platelet-derived growth factor receptor β
$pH_{1/2}$	pH of mid transition
PNA	peptide nucleic acid
$POCl_3$	phosphoryl chloride
PS	phosphorothioate
q	quartet
R	ideal gas constant
RAP	repeat addition processivity
RET	rearranged during transfection
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNAi	RNA interference
RRMs	RNA recognition motifs
r.t.	room temperature
ΔS	entropy
s	singlet
shRNA	short hairpin RNA
siRNA	short interfering RNA
ss	single-stranded
SWNTs	single-walled carbon nanotubes
T	thymine
t	triplet
$T_{1/2}$	melting temperature when the system is not at equilibrium
T-loop	telomeric loop
TAE	tris-acetate-EDTA buffer
TCA	trichloroacetic acid
Tc-DNA	tricyclo-DNA
TERRA	telomeric repeat-containing RNA

THF	tetrahydrofuran
TLR	toll-like receptor
T_m	melting temperature at equilibrium
TMSCl	trimethylsilyl chloride
TOCSY	total correlation spectroscopy
TRAP	telomere repeat amplification protocol
TREAT-HF	triethylamine trihydrofluoride
TSS	transcriptional start sites
U	uridine
UNA	unlocked nucleic acid
UTR	untranslated region
UV	ultraviolet
WC	Watson-Crick
VEGF	vascular endothelial growth factor

Chapter 1

Introduction



"You don't have to be great to start,
but you have to start to be great."
–Zig Ziglar

1.1 Nucleic Acids: The Building Blocks of Life and Beyond

"Rather than believe that Watson and Crick made the DNA structure, I would rather stress that the structure made Watson and Crick." –Francis Crick

The discovery of nucleic acids, the fundamental molecules of life, sparked the interest of talented researchers in further investigating their structure and determining their various functions. Indeed, studying nucleic acids is of paramount importance due to their role as the carriers of genetic information. Furthermore, nucleic acids are being implemented in several research areas ranging from biology, chemistry and medicine to diagnostics, forensics, and nanotechnology.

The field of nucleic acid chemistry has benefitted tremendously from the use of chemical modifications that facilitated structural determination and design of nucleic acid therapeutics. From this perspective, the aim of this thesis is to incorporate chemically-modified nucleic acids in telomeric C- and G-rich sequences in order to stabilize their desired folding topologies, and determine their structure and function in telomere biology.

1.2 Nucleic Acid Structure and Function

The elucidation of the helical structure of double-stranded deoxyribonucleic acid (DNA) was a result of numerous key discoveries by talented scientists over several decades. In 1868, Friedrich Miescher isolated and precipitated cellular nuclei, which he obtained from pus cells on discarded bandages, and named the phosphorus-containing material nuclein.[1] Miescher also found this material to be strongly acidic. In 1889, Richard Altman further isolated nuclein from its proteinaceous component and called the substance nucleic acid.[1] In 1923, Fred Griffith made a key discovery in bacterial genomics. He showed that the hereditary change from the non-virulent form to the virulent form in pneumonia bacterium is permanent.[2] The four deoxyribonucleotides of DNA were discovered by Klein and Thannhauser in 1935, upon the discovery that DNA could be enzymatically cut into four different mononucleotides. Oswald Avery's discovery in 1940 was a breakthrough in the field of nucleic acid chemistry. He showed that DNA was responsible for the bacterial transformation observed by Fred Griffith.[3] In 1944 Alexander R. Todd established the chemical synthesis of the DNA nucleotides for which he was awarded the Nobel Prize in Chemistry in 1957. Later, Erwin Chargaff showed that different

species have different nucleobase composition of DNA in their genome, but that the ratio of purines is always equal to the ratio of pyrimidines.[4] This observation laid the groundwork for determining the structure of the DNA double helix. Maurice Wilkins and Rosalind Franklin joined forces to determine the X-ray diffraction pattern for DNA. They observed that DNA adopts two highly crystalline helical forms: an A-form at low humidity and high salt conditions, and a B-form at high humidity and low salt conditions.[5,6] Franklin concluded that the phosphate groups must be exposed to water, *i.e.* on the outside of the helix.

James Watson and Francis Crick combined the pieces of information from previous researchers and proposed several models before they established the double helical structure of DNA in 1953.[7] Applying Chargaff's base-pairing rule, it became clear to Watson and Crick that the nucleobase sequence of one strand automatically dictated the sequence of the second strand. Therefore, one strand can be used as a template for copying the genetic material.[8] In 1962, James Watson and Francis Crick were awarded the Nobel Prize in physiology and medicine, along with Maurice Wilkins for discovering the molecular structure of nucleic acids and determining their role in transferring genetic information. The unique features of nucleic acids, such as data-storage and information-processing capacity, highly specific binding properties, and particular hybridization ability, were beneficial to several fields of research.

1.2.1 Natural Nucleosides and Nucleotides

Nucleic acids are long biopolymers composed of a linear array of monomeric units termed nucleotides.[1] As shown in **Figure 1.1**, a nucleotide is a phosphorylated nucleoside composed of a pentose sugar, a nitrogenous heterocyclic nucleobase and a phosphate group. The two major classes of nucleic acids are ribonucleic acids (RNA) and deoxyribonucleic acids (DNA); each possesses a slightly different structure. RNA contains D-ribose sugar while the DNA sugar lacks the 2'-OH and is therefore termed 2'-deoxy-D-ribose. The nucleobases are attached to the anomeric carbon (C1') of the sugar moiety *via* a β-glycosidic linkage. There are two types of nucleobases: monocyclic pyrimidines [cytosine (C), thymine (T) and uracil (U)] and bicyclic purines [adenine (A) and guanine (G)]. DNA and RNA share the same purines and pyrimidines except that RNA contains uracil while DNA replaces uracil with thymine. In an oligonucleotide (ON), nucleotides are linked consecutively *via* phosphodiester linkages between

the 3'-hydroxyl group of one nucleotide and the 5'-hydroxyl group of the next nucleotide. Furthermore, ON sequences are written from the 5'-end to the 3'-end by convention (**Figure 1.1**).

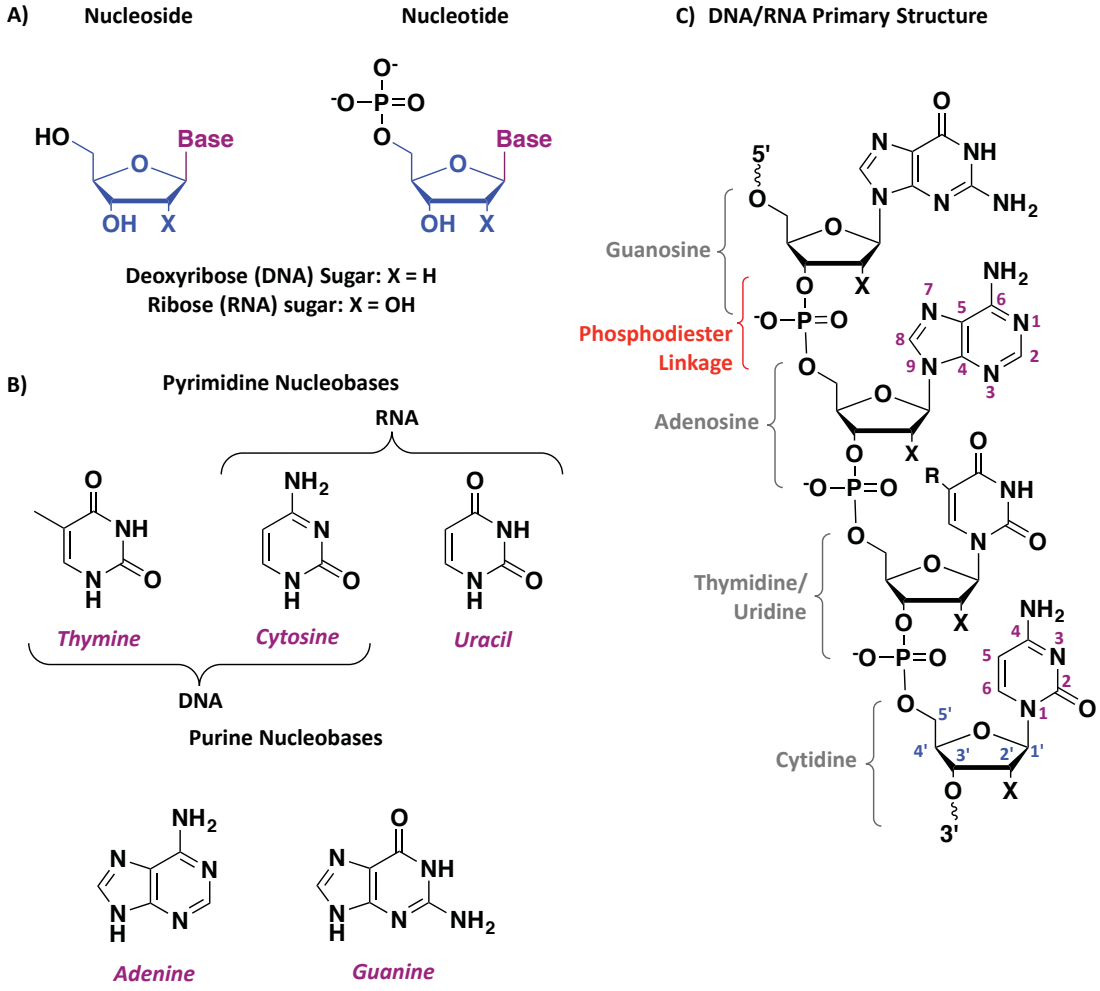


Figure 1.1: Nucleic acids components. A) Nucleosides consist of a nucleobase and a 2'-deoxyribose sugar in case of DNA and a ribose sugar in RNA. The nucleobases are attached to the pentose sugar at the 1' carbon *via* a β-linkage. Nucleotides are phosphorylated nucleosides. B) The pyrimidine and purine nitrogenous heterocyclic nucleobases. RNA contains uracil while DNA contains thymine. C) Example of a DNA/RNA primary structure. Adenine and cytosine as well as one of the sugars are numbered following the convention. In Thymidine R = CH_3 and in Uridine R = H.

1.2.2 Conformation of Nucleosides and Nucleotides

Furanose sugar rings are inherently nonplanar with some of their atoms twisted out of plane to relieve strain; this twist, or nonplanarity, is called puckering. The sugar pucker is identified based on the orientation of carbon-2' and carbon-3' with respect to the plane of C1'-

O4'-C4' (**Figure 1.2**). The term *endo* implies that ring atoms are on the same side as the C5' and the nucleobase, while the term *exo* implies that they are on opposite sides (**Figure 1.2**).

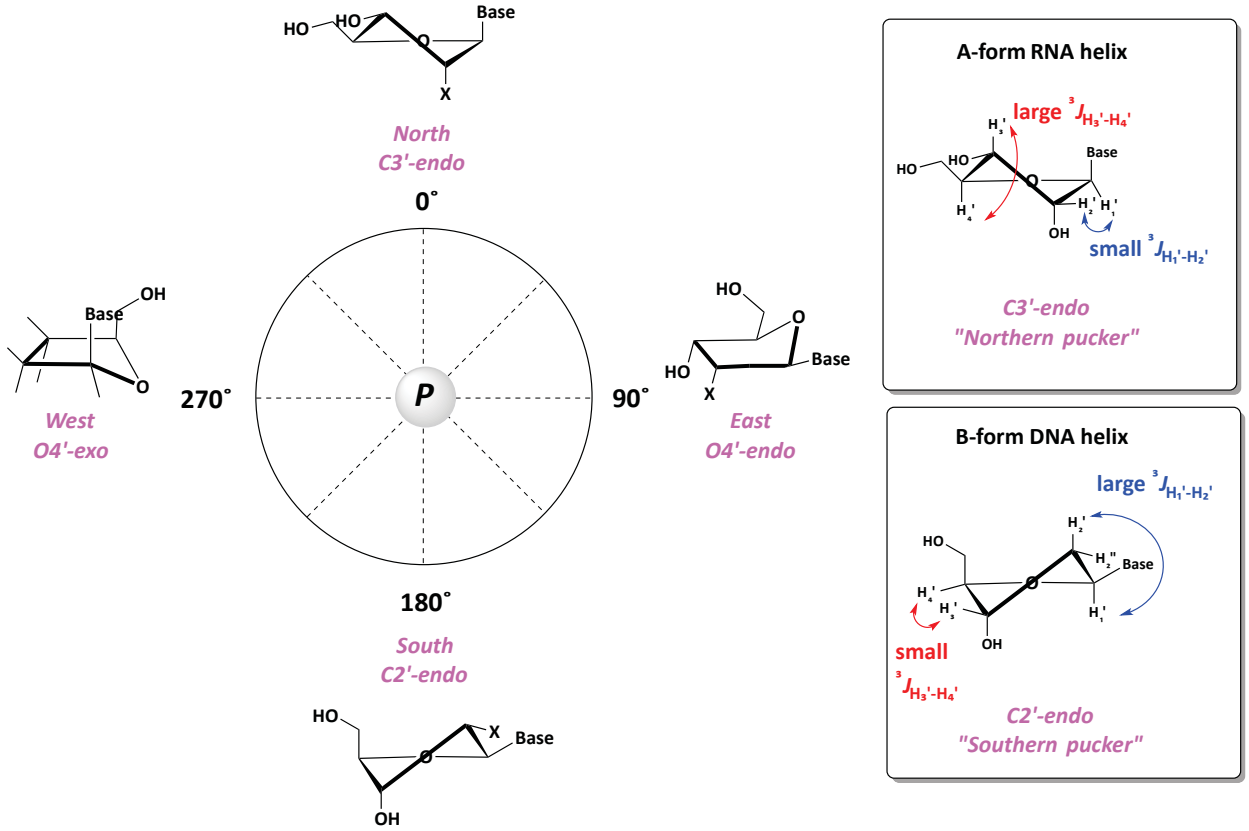


Figure 1.2: The pseudorotational wheel and sugar pucker notations of the furanose ring. Nucleoside sugar puckers are classified based on the phase angle (*P*). In A-form helices, the sugar adopts a North (C3'-*endo*) conformation while B-form DNA sugars adopt a South (C2'-*endo*) conformation.

In 1972, Altona and Sundaralingam introduced the concept of pseudorotation, to more precisely determine the sugar pucker conformation based on two parameters: the phase angle of pseudorotation (*P*) and the degree of pucker (τ_m). These parameters are determined from the five endocyclic torsion angles. Sugar puckering conformations, based on phase angle values, are commonly represented on a pseudorotational wheel (**Figure 1.2**).[9] In solution, nucleosides and nucleotides equilibrate rapidly between different conformations; these conformational minima are characterized on the pseudorotational wheel as either North (C3'-*endo*, *P* in the range of -10° to + 40°) or South (C2'-*endo*, *P* in the range of 144-180°) passing through East (O4'-*endo,* **Figure 1.2**). The nature of the substituents on the sugar ring shifts the equilibrium and populates a specific region of the pseudorotational wheel due to a combination of anomeric and gauche

effects. For example, 2'-deoxyribonucleosides mainly adopt a South conformation (65%), whereas the ratio of North to South in ribonucleosides is around 51% to 49%.[10]

^{1}H-NMR data can be utilized to unambiguously determine the sugar conformation of nucleosides/nucleotides. Sugar pucker geometries directly influence the three bond ^{1}H-^{1}H scalar coupling within the sugar moiety, in particular $^{3}J_{H'-H2'}$ and $^{3}J_{H3'-H4'}$ coupling constants (**Figure 1.2**). These coupling constants can be used to resolve the North/South conformational distribution of a given nucleoside in solution.[11,12] For instance, the South (C2'-*endo*) conformation is characterized by a large $^{3}J_{H'-H2'}$ coupling constant and a small $^{3}J_{H3'-H4'}$, while the North conformation is characterized by a large $^{3}J_{H3'-H4'}$ and a small $^{3}J_{H'-H2'}$ coupling constant (**Figure 1.2**). For an optimal North conformation, $P = 0°$ and the coupling constant $^{3}J_{H'-H2'}$ is also equal to zero.

1.2.3 The DNA Double Helix

Watson and Crick's discovery of the DNA double helix, also known as DNA secondary structure, was a major breakthrough due to its biological implications.[7] In the proposed model, the two strands of the duplex are antiparallel, the negatively charged phosphodiester backbones are on the exterior, and the hydrophobic nucleobases are on the interior of the helix stacked above one another. The two strands are held together by intermolecular noncovalent hydrogen bonds between nucleobases in an arrangement termed "Watson-Crick" (WC) base-pairing. Adenine forms two hydrogen bonds with thymine while guanine forms three hydrogen bonds with cytosine (**Figure 1.3**). In addition to base-pairing, nucleic acid secondary structures are stabilized by π-π base-stacking interactions between sequential nucleobases.[13-18] However, the duplex is also destabilized by electrostatic repulsion between the negatively charged phosphate residues in the backbone.[19] Therefore, the thermodynamic stability of nucleic acids is an average of stabilizing and nonstabilizing factors.

Double-stranded DNA (dsDNA) typically adopts a B-form helical conformation and is associated with South (C2'-*endo*) sugar puckers,[20,21] while dsRNA adopts an A-form conformation and is associated with North (C3'-*endo*) sugar puckers.[22,23] Due to the flexible nature of 2'-deoxyribose, DNA can adopt several conformations depending on the nucleobase sequence, the solvent, and the presence of small molecule ligands or metal ions.[24,25] For example,

as observed by Rosalind Franklin, DNA duplexes can also adopt an A-form conformation in conditions of low humidity.[6,26]

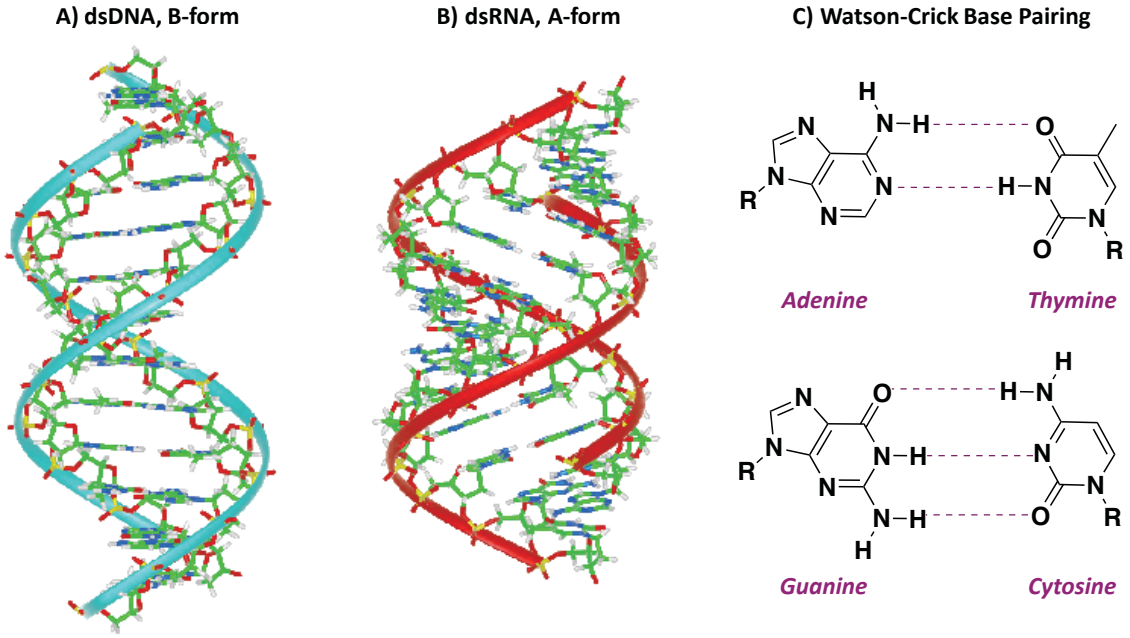


Figure 1.3: A) B-form double-stranded DNA. B) A-form double-stranded RNA. C) Schematic representation of the canonical Watson-Crick base-pairing. Adenine (A) forms two hydrogen bonds with thymine (T) and guanine forms three hydrogen bonds with cytosine (C).

A-form and B-form DNA are antiparallel, right-handed double helices with the sugar-phosphate backbone parallel to the helix axis; however, they possess several conformational differences, some of which are highlighted in **Table 1.1**. B-form DNA helices are flexible and possess a wide major groove and a narrow minor groove, while A-form RNA duplexes are compact and feature a deep major groove and a wide minor groove. When DNA is bound to an RNA strand, the hybrid duplex will mainly adopt an A-form conformation since the South sugar conformation in the more flexible DNA strand shifts to North to form a stable duplex with RNA.[27] Today, nucleic acid structural features are determined at atomic resolution *via* X-ray crystallography and NMR spectroscopy. It is crucial to know the parameters of the major and minor grooves of DNA and RNA double helices since proteins and small molecules interact with nucleic acids through these grooves.[28,29]

Table 1.1: Average helical parameters for B-form dsDNA and A-form dsRNA.[1]

Helical Parameters	B-form	A-form
Helix Sense	Right-handed	Right-handed
Sugar Pucker	South (C2'-*endo*)	North (C3'-*endo*)
Nucleotides per turn	10	11
Major Groove Width	~11.7 Å	~2.7 Å
Major Groove Depth	~8.8 Å	~13.5 Å
Minor Groove Width	~5.7 Å	~11 Å
Minor Groove Depth	~7.5 Å	~2.8 Å

1.2.4 The *Central Dogma*: The Mechanism of Gene Expression

Following the elucidation of the DNA double helix structure, the main question to tackle was: *how is the genetic code transferred from a DNA sequence into proteins*? The *central dogma* of molecular biology illustrates the transition from DNA to proteins responsible for cellular functions. Genomic DNA, which stores genetic information, is found in the nucleus and surrounded by proteins. During DNA replication, DNA polymerase copies a DNA strand into two identical daughter strands.[30] The synthesis of RNA from genomic DNA is carried out by RNA polymerases.[31] First, DNA is converted into single-stranded RNAs (precursor mRNAs), which are longer than the desired mature product, in a process called *transcription*. The nascent pre-mRNA undergoes maturations *i.e.* post-transcriptional modifications such as splicing during which the noncoding intronic regions are removed and the coding exon regions are joined in the nucleus to generate the mature messenger RNA (mRNA).[32-34] mRNA, an intermediary species containing the sequence of genes copied from DNA,[35] is then exported from the nucleus and transported to ribosomes in the cytoplasm, where it serves as a template for protein biosynthesis, in a process termed *translation*.[36] Ribosomes are composed of proteins and catalytic RNAs,[37,38] called ribosomal RNA (rRNA), which utilize highly-ordered transfer RNA (tRNA) to synthesize functional proteins from amino acids.[39] In sum, the biosynthesis of proteins is dictated by the nucleobase sequence of DNA, where all the genetic information of the diverse biological molecules is stored.

1.3 Chemically-Modified Nucleosides and Oligonucleotides

Chemically-modified nucleoside analogues as antiviral and anticancer agents and modified ONs as gene silencing therapeutics continue to spark interest in the field of nucleic acid chemistry. Several modified nucleosides and ONs have been approved by the FDA for the treatment of a wide range of diseases; however, demand persists for modifications with improved potency and specificity along with lower toxicity and drug-development cost. Resistance to previously approved drugs and the emergence of new diseases necessitate the development of novel therapeutic agents. The ON therapeutics field has benefited greatly from the chemically-modified nucleoside/nucleotide analogues where sugar puckering, nuclease stability, binding affinity to target mRNA, and immunostimulatory profiles are controlled, among other factors.

1.3.1 Modified Nucleosides as Therapeutic Agents

The study of nucleoside analogues dates back several decades and has led to the development of antiviral and anticancer compounds. To date, twenty nucleoside analogues have been approved for the treatment of cancer, HIV (human immunodeficiency virus), HSV (herpes simplex virus), HBV (hepatitis B), HCV (hepatitis C), and HCMV (human cytomegalovirus).[40-45] These analogues are structurally similar to the natural deoxynucleoside triphosphates (dNTPs) used during DNA polymerization; however, they lack the 3'-hydroxyl group on the sugar ring. Once inside the cell, the nucleoside analogues must undergo stepwise phosphorylation by cellular kinases to yield the mono-, di- and tri-phosphorylated active analogue (**Figure 1.4**).[46,47]

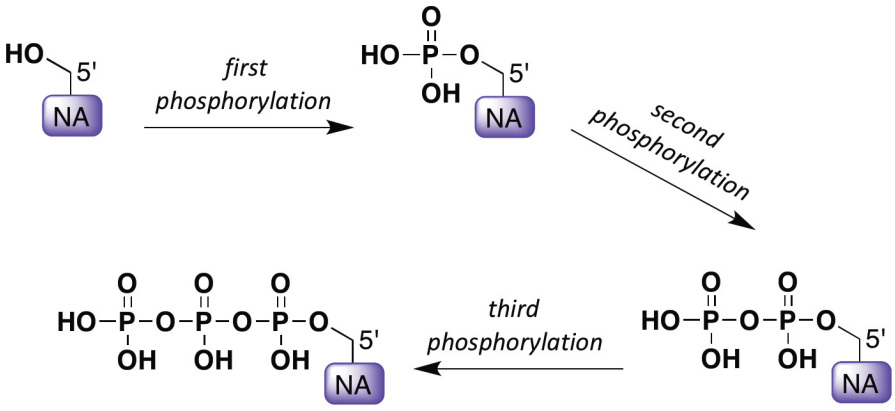


Figure 1.4: Activation of nucleoside analogues (NAs) into their 5'-triphosphorylated form.

Nucleoside analogues target and inhibit intracellular enzymes or lead to the chain termination of the viral nucleic acids.[44,45,48] The activation of these analogues, even those possessing very similar structures to dNTPs, proceeds inefficiently due to the substrate specificity of kinases. Nucleoside conformation plays a major role in recognition by kinases and polymerases: kinases require nucleosides that adopt a southern sugar pucker,[49] while polymerases incorporate nucleosides that adopt a northern sugar pucker.[50,51] These modified nucleotides should be substrates for viral polymerase and kinases and not for non-infected host cells: for instance, the low fidelity of reverse transcriptase and lack of proofreading ability facilitates the incorporation of modified dNTPs.

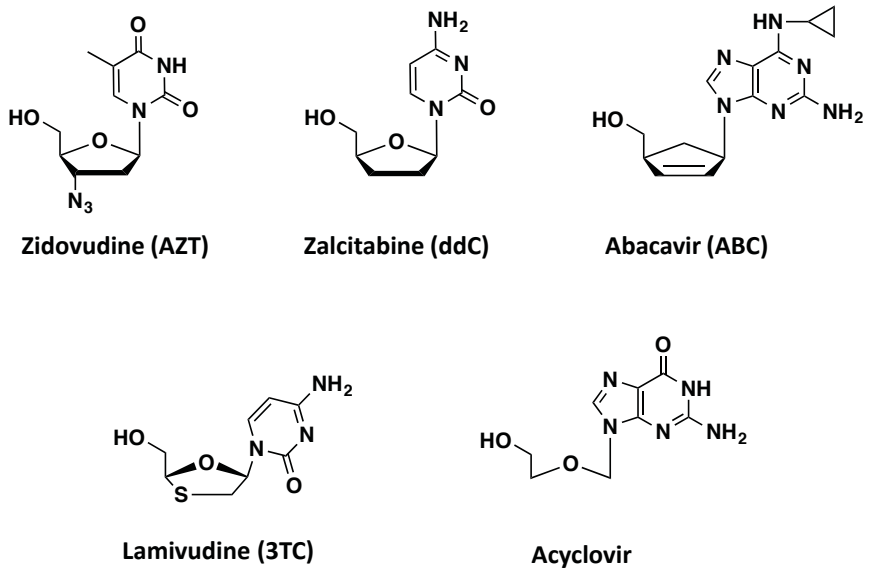


Figure 1.5: Structures of some FDA approved nucleoside analogues.

Typically, nucleoside analogues are designed with modifications within their sugar or nucleobase moiety. Some key examples are listed in **Figure 1.5**.[44,52] AZT (3'-azido-2',3'-dideoxythymidine) was the first 2',3'-dideoxynucleoside reverse transcriptase inhibitor (NRTI) approved for the treatment of HIV in 1987;[53] it possesses a 3'-azido group instead of the 3'-hydoxyl group. The next two compounds approved for HIV treatment were 2',3'-dideoxyinosine (ddI) and 2',3'-dideoxycytidine (ddC) where the 3'-hydroxyl group was removed.[54] Other analogues contain an alkene in the sugar moiety such as 2',3'-didehydro-2',3'-dideoxythymidine (d4T) and abacavir (ABC); ABC also possesses a carbocyclic sugar ring.[55,56] Lamivudine (3TC, (-)-2',3'-dideoxy-3'-thiacytidine) was discovered at McGill University by Bernard Belleau, who

pioneered the introduction of heteroatoms within the furanose ring. 3TC was FDA approved in 1995 for the treatment of HIV.[57] Emtricitabine [(–)FTC], a 5-fluorocytosine analogue of 3TC, was approved in 2003.[58] An acyclic nucleoside analogue, Acyclovir, was discovered by Gertrude B. Elion for the treatment of cold sores, shingles and herpes genitalis; for this work, Elion was awarded the Nobel Prize in Physiology or Medicine in 1988.[59]

Despite the wealth of modified analogues synthesized so far, there is still a demand for new nucleoside analogues, and their respective prodrugs, to overcome several limitations of currently available therapeutics, such as inefficient phosphorylation, poor oral bioavailability, and drug resistance.[60] Resistance occurs *via* several mechanisms such as: discrimination and nucleotide excision,[61] inactivation of nucleoside analogues due to catabolic bioconversion,[62,63] and dephosphorylation of the active 5'-triphosphorylated analogues by 5'-nucleotidases.[64,65] The use of highly active antiretroviral therapy (HAART) has helped diminish drug resistance. In order to bypass the rate-limiting mono-phosphorylation step,[66,67] mono-phosphorylated (monophosphonate) prodrugs possessing P-CH_2- bonds, rather than P-O- bonds, have been prepared. This approach improved their *in vivo* stability and transport into targeted cells,[68] and therefore led to the development of three clinically-approved antiviral compounds with longer half-lives and improved *in situ* stability: cidofovir for the treatment of HCMV, adefovir for the treatment of HBV, and tenofovir for the treatment of HIV.[44] Additional prodrug approaches were developed to improve the cellular uptake of the previously mentioned monophosphonates.[69] For example, Chris McGuigan pioneered the ProTide prodrug technology, in which the hydroxyl groups of the monophosphate or monophosphonate analogues are masked by aromatic or amino acid ester groups, rendering the molecules more lipophilic and improving their delivery into the cells.[70] The ProTide approach has led to two FDA-approved antiviral compounds: tenofovir alafenamide (TAF), and sofosbuvir (**Figure 1.6**),[71] as well as several candidates in clinical trials for the treatment of HIV, HBV, HCV, Ebola, and cancer.[72] Even after surpassing the first phosphorylation step, the di- and tri- phosphorylation steps can also create a bottleneck that renders the metabolic conversion of nucleoside analogues into their active triphosphate forms inefficient. For instance, the di-phosphorylation step of AZT by thymidylate kinase (TMP-K) is rate-determining;[73] similarly, the conversion of d4U or ddU into tri-phosphorylated forms by nucleoside diphosphate kinase (NDP-K) proceeds inefficiently.[74] In order to bypass the

phosphorylation steps, Chris Meier's group has successfully developed nucleoside diphosphate (Di*PP*ro) and nucleoside triphosphate (Tri*PPP*ro) delivery systems.[75,76]

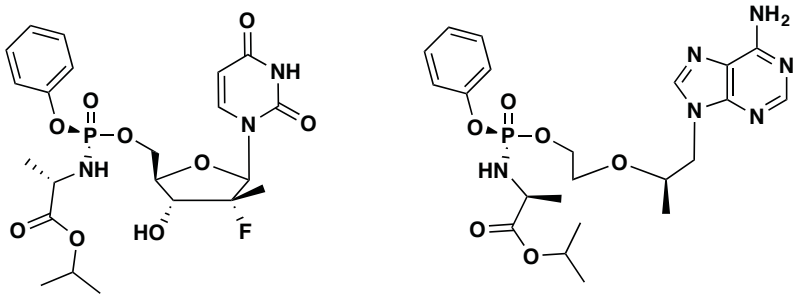


Figure 1.6: Chemical structures of sofosbuvir and tenofovir alafenamide (TAF), the two FDA approved antiviral ProTides.

1.3.2 Oligonucleotides as Gene Silencing Therapeutics

The concept of utilizing ONs as gene-silencing therapeutics dates back to 1978 when Zamecnik and Stephenson reported the use of AONs to inhibit viral replication.[77,78] Since then, the field has expanded greatly, and new frontiers have emerged with the latest breakthrough discovery of the CRISPR / Cas9 (clustered regularly interspaced short palindromic repeats) technology and its application in genome editing.[79] ON therapeutics act at the post-transcriptional level and have been categorized, based on their mechanism of action, into the following classes: antisense oligonucleotides (AONs),[77,78,80] short interfering RNAs (siRNAs),[81] microRNA-targeting ON (anti-miRNAs),[82,83] exon-skipping and splice-switching compounds,[84] immunostimulatory analogues,[85] and DNA/RNAzymes.[86,87] A detailed description of these compounds and their mechanism of action is beyond the scope of this thesis, and the reader is directed to extensive reviews covering this subject.[88-90] However, a brief description will be provided for AONs and siRNAs.

Antisense-mediated gene targeting paved the way for ON-based therapeutics. AONs that target a specific region in the mRNA of a gene can modulate gene expression through two distinct mechanisms; a steric block approach and an RNase H-dependent approach.[91] AONs in the steric block approach are designed with a very strong binding affinity to the target mRNA, preventing ribosomal assembly and leading to translational inhibition without cleaving the

mRNA.[91] On the other hand, in the RNase H-dependent mechanism, AONs bind to the target mRNA and recruit the RNase H enzyme, leading to the cleavage of the mRNA target.[92] The RNase H enzyme is an endogenous endonuclease known to cleave the RNA strand of a DNA:RNA hybrid duplex.[91,93]

RNA inference pathway (RNAi) can be triggered either by an endogenous 21-nt long double-stranded RNA known as microRNA (miRNA),[94] or by an exogenous dsRNA termed siRNA, designed to mimic miRNAs.[95] siRNAs are typically 21 to 24-nt long with 2-nt overhangs at the 3'-end. The guide (antisense) strand is loaded into the RNA-induced silencing complex (RISC) and is complementary to the mRNA target, while the passenger (sense) strand is discarded. The extent of complementarity between the miRNA or siRNA guide strand and the target mRNA dictates whether gene silencing occurs through translational arrest or mRNA cleavage.[94]

While there has been great optimism surrounding the development of AON- and RNAi-based therapeutics, these approaches suffer from several limitations, such as poor biostability (particularly RNA) due to degradation by intracellular nucleases and serum,[88] and the ineffective delivery to targeted cells.[96] Furthermore, off-target effects (OTEs) could occur when unintended mRNAs are knocked or when the AONs and siRNAs trigger an immune response by interactions with cellular immune receptors.[97,98] In order to bypass these limitations, delivery vehicles[99,100] and several modifications (nucleobase, sugar and/or internucleotide linkage) have been utilized.[89] A list of chemical modifications utilized in ON therapeutics is shown in **Figure 1.7**.

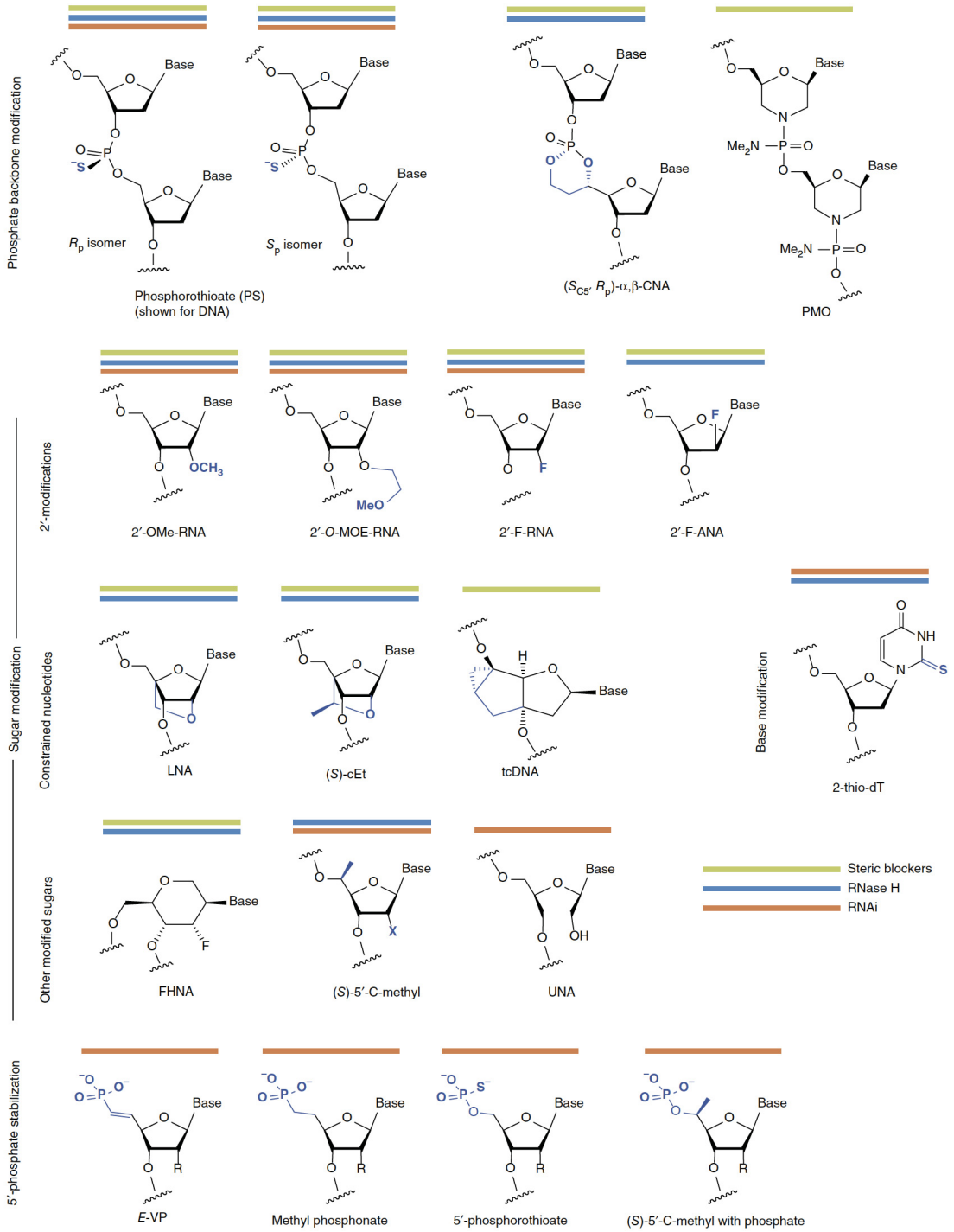


Figure 1.7: Chemical structures of some of the sugar, nucleobase, and internucleotide linkage modifications utilized in ON therapeutics. The colored lines indicate the different mechanism of action of ONs containing these modifications. Adapted with permission from reference [90] (Nature Publishing Group, 2017).

The first AON-based drug was approved by the FDA for the treatment of cytomegalovirus retinitis in 1998. This AON, named Vitravene (fomivirsen), consists of a phosphorothioate backbone (PS) (**Figure 1.7**).[90] The second approved AON drug was Kynamro (mipomersen) for the treatment of hypercholesterolemia. Kynamro has a combination of internucleotide linkage modifications (PS) and sugar modifications (2'-*O*-methoxyethyl (2'-*O*-MOE), **Figure 1.7**).[101] Macugen, an aptamer drug, was approved by the FDA in 2004 for treating age-related macular degeneration (AMD).[102] 2'-F and 2'-*O*-methyl (2'-*O*-Me, **Figure 1.7**) modifications are utilized in Macugen in order to preorganize the sugars into the C3'-*endo* conformation. Locked nucleic acids (LNA)[103] and Tricyclo-DNA (tc-DNA, **Figure 1.7**)[104,105] are constrained nucleic acids that exhibit high binding affinity towards RNA due to the reduced conformational flexibility.[106] Recently two splice-switching ONs have been approved by the FDA, Exondys 51 and Spinraza. Exondys 51 (eteplirsen) is a 30-nt long phosphorodiamidate morpholino oligomer (PMO, **Figure 1.7**) used for treating Duchenne muscular dystrophy,[107] whereas Spinraza (nusinersen) is a fully modified MOE 18-mer AON used in the treatment of spinal muscular atrophy.[108] The field of ON therapeutics holds great promise and the use of chemical modifications and delivery vehicles could place ONs at the forefront of drug discovery, together with small molecules and monoclonal antibodies.[90]

1.3.3 2'-Deoxy-2'-Fluoro-β-D-Arabinonucleic Acid (2'F-ANA)

2'-Deoxy-2'-fluoro-β-D-arabinonucleic acid (2'F-ANA) modifications, in particular 2'-arabinofluorocytidine and 2'-arabinofluoroguanosine, are two nucleoside analogues utilized in the research described in this thesis. 2'F-ANA is the 2'-epimer of 2'F-RNA (**Figure 1.8**); however, due to the gauche effects caused by the fluorine atom in the arabino position, the nucleoside tends to adopt a South-East conformation, making it an excellent structural mimic of DNA.[109,110]

2'F-ANA modified ONs exhibit great potential as therapeutics and in structural studies of non-canonical DNA structures, such as G-quadruplexes (G4s) and i-motifs, both of which are explored in great detail in this thesis. 2'F-ANA was found to enhance binding affinity to complementary RNA and improve stability against exo- and endonucleases in siRNA and AON-based drug candidates.[111-113] Of note, 2'F-ANA is among the few sugar-modified ONs that elicit RNase H-mediated cleavage of target mRNAs.[114] Chimeric siRNA constructs that combine 2'F-

ANA with RNA-like modifications such as LNA and 2'F-RNA exhibit enhanced siRNA potency and reduced immunostimulation profiles.[112] The stability observed in 2'F-ANA:RNA duplexes is due to C-H···F-C pseudohydrogen bonding between the 2'-fluorine and purine H8 at pyrimidine-purine steps.[115] Moreover, Clofarabine, a 2'F-ANA nucleoside analogue, is FDA approved for the treatment of leukemia (**Figure 1.8**).[116]

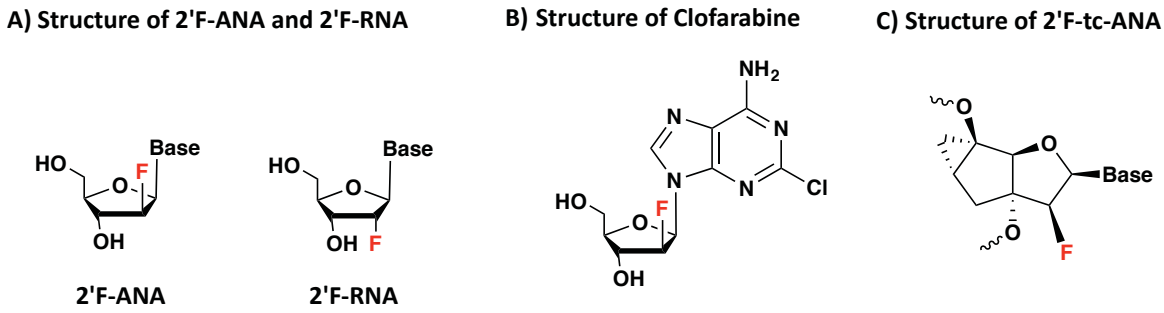


Figure 1.8: A) Chemical structure of 2'F-ANA and its 2'-epimer, 2'F-RNA. B) Structure of Clofarabine, a 2'F-ANA nucleoside analogue FDA-approved for the treatment of leukemia. C) Structure of 2'β-fluoro-tricyclo nucleic acids (2'F-tc-ANA).

Additionally, our group was able to synthesize the first microarrays based on 2'F-ANA and 2'F-ANA/DNA chimeric sequences,[117] and utilized them to identify aptamers with high binding affinity to thrombin. Very recently, the Leumann group synthesized a new 2'F-ANA analogue, 2'β-fluoro-tricyclo nucleic acid (2'F-tc-ANA, **Figure 1.8**).[118] This modification exhibits high binding affinity when paired with DNA (+1.6 °C/mod) and RNA (+2.5 °C/mod) and promotes RNase H-mediated cleavage of the complementary RNA. Therefore, 2'F-tc-ANA is a promising modification for ON-based therapeutics.

1.4 Beyond the DNA Double Helix and Watson-Crick Base-Pairing

Since the discovery of the B-DNA double helix,[7] it has been shown that DNA adopts several polymorphs.[119] During replication and transcription processes, double-stranded DNA partially unwinds into single-stranded sequences.[120] It has been reported that more than 50% of the human genome is comprised of repetitive DNA sequences.[121] Therefore, under certain conditions, single-stranded repetitive DNA strands fold inter- or intramolecularly into unique non-B-form secondary structures such as left-handed Z-DNA,[122] poly (dA) duplexes (A-motif),[123] hairpins,[124] triplexes,[125] G4s,[126] and i-motifs.[127]

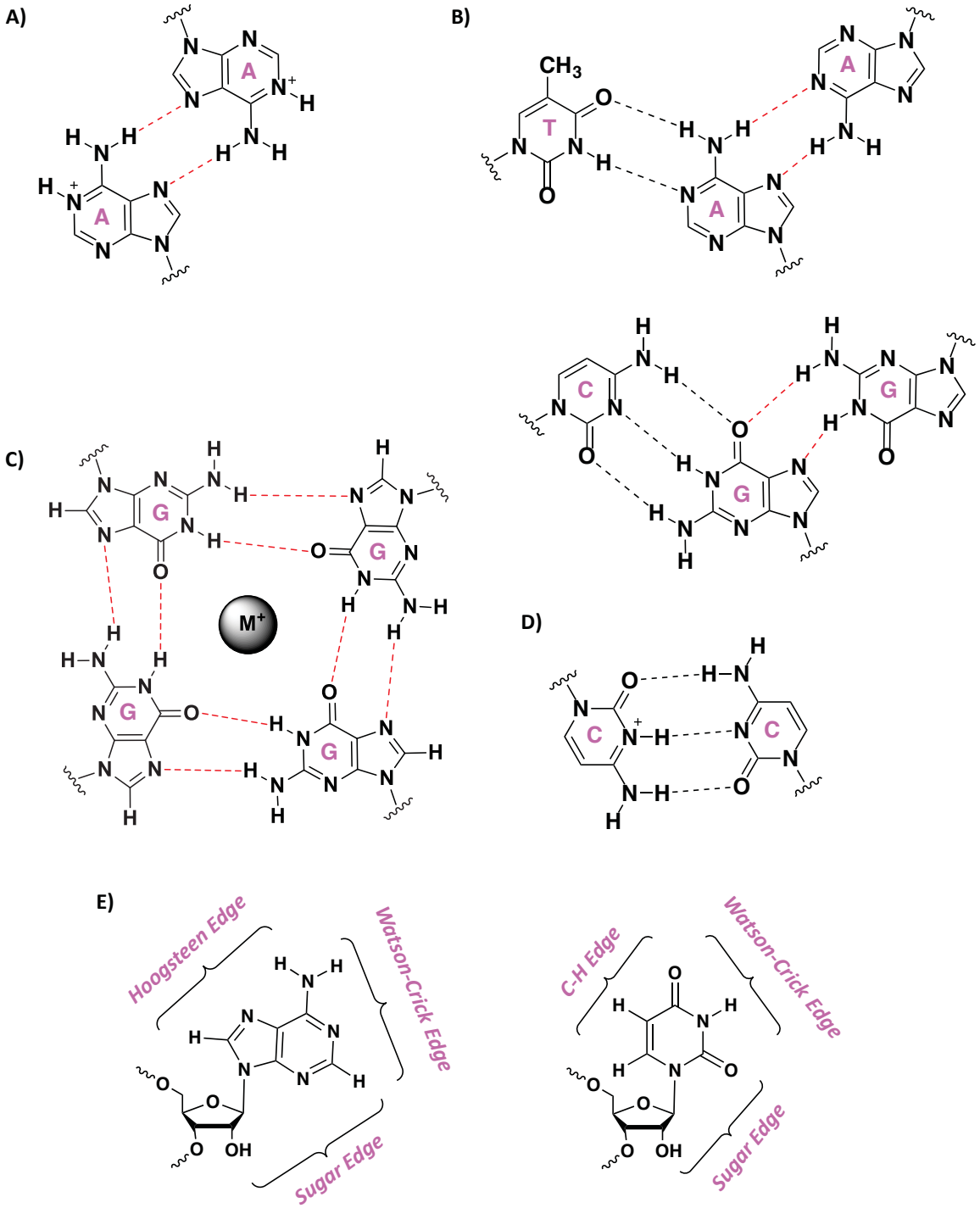


Figure 1.9: Some non-canonical hydrogen-bonding interactions in: A) A-motifs, B) tetraplexes, C) G-tetrad with a central cation, and D) i-motifs. WC and reverse WC base pairs are shown in black dashes, Hoogsteen and reverse Hoogsteen are shown in red dashes. E) Scheme depicting the different hydrogen bonding edges.

There is increasing evidence for the existence of these structures in cells and their potential role in controlling gene expression and replication.[128,129] Moreover, they have been attracting considerable research interest in the pharmaceutical and medical fields due to their significant roles in genetic instability and human disease.[130]

A free database for locating sequences with the tendency to fold into non-canonical secondary structures across the genome has been created by Stephens and coworkers.[131] Among these structures are A-motif duplexes, triplexes, and G4 and i-motif tetraplexes held together by "unusual" base pairs and base triplets as shown in **Figure 1.9**.[120] The canonical antiparallel double helical DNA structure is stabilized by WC base-pairing where the two bases interact through their WC edges (**Figure 1.9**) with *cis* glycosidic bond orientation (relative to the axis of rotation).[132] X-ray crystallography and high resolution nuclear magnetic resonance (NMR) revealed that hydrogen-bonding can be divided into three categories/edges: the Watson-Crick edge (WC) present in regular B-DNA double helices; the C-H, or Hoogsteen edge (H), which stabilizes G4 structures; and the sugar edge (SE) which is relevant to RNA where the 2'-hydroxyl group in the ribose sugar is available for hydrogen bonding (**Figure 1.9**).[132]

In this work, we are particularly interested in stabilizing i-motif structures at physiological conditions (Chapters 2 and 3) and investigating the effect of the co-existence of i-motifs and G4s in human telomeric sequences on telomerase activity (Chapter 4).

1.5 Human Telomere Biology: Telomeres and Telomerase in Cellular Aging and Cancer

1.5.1 Telomeres: The Protective Caps of Chromosomes

In 1938, Hermann Muller was the first to discover that the tip of chromosomes possesses special structural features (**Figure 1.10**). Based on their position at the end of chromosomes, Muller named these ends "telomeres" (from Greek *telos* 'end' + *mere* 'part').[133] Barbara McClintock later observed that chromosomes with broken ends tend to fuse together by DNA repair processes, hence suggesting that telomeres are necessary to protect chromosomes from these fusion events. These broken ends result from the "end replication problem," proposed by James Watson in 1972.[134]

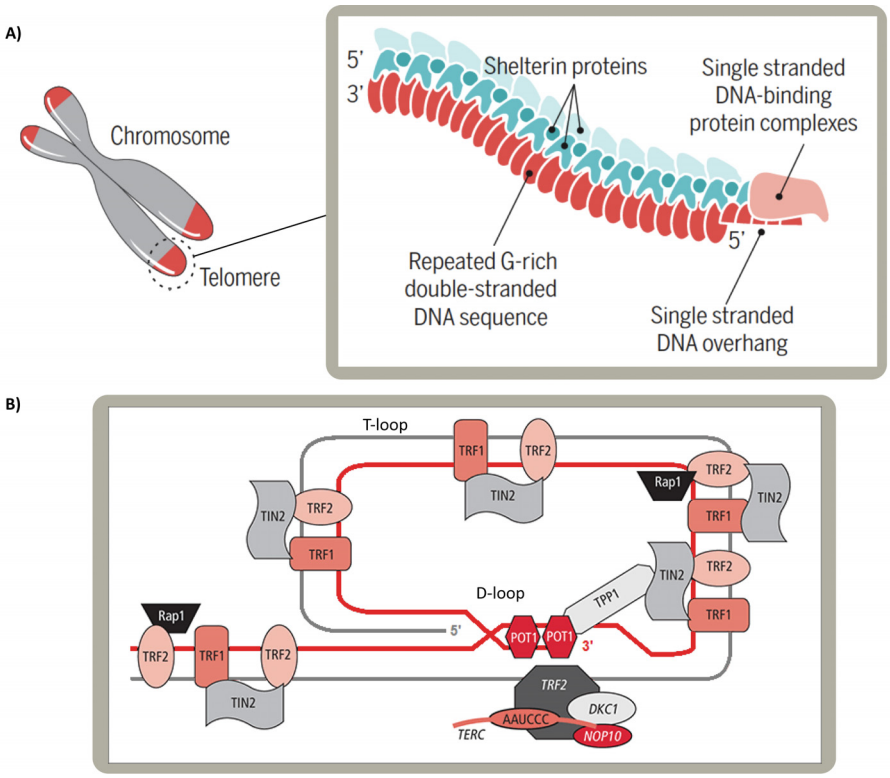


Figure 1.10: A) Representation of a chromosome and of the telomeric region. Figure has been modified from reference [135a]. B) Scheme representing the association of the telomerase and the shelterin component proteins to telomere ends. The shelterin complex is composed of six proteins: TRF1, TRF2, POT1, TIN2, TPP1, and Rap1. T-loop stands for "telomeric loop" and D-loop stands for "displacement loop". Figure modified from reference [135b].

During DNA replication one of the two daughter strands, the leading strand, is synthesized continuously, while the other strand, the lagging strand, is synthesized in short Okazaki fragments. The DNA repair machinery fills the gaps between these fragments; however, DNA polymerase is unable to add nucleotides at the 3'-end due to the absence of the RNA primer. Therefore, the daughter strands will always be at least one RNA primer shorter than the parent DNA strand, and chromosomes will shorten with every cell-cycle division. Essentially this is a cause of aging and limited cellular lifespan.

It was not until 1978 -forty years after Muller's initial discovery- that Elizabeth Blackburn observed one repeat sequence $(TTGGGG)_n$ in chromosomal ends of *Tetrahymena* (a unicellular eukaryotic ciliate).[136] Likewise, human telomeres consist of highly repetitive hexameric repeats bound by the component proteins of the shelterin complex (**Figure 1.10**).[135] The hexamer telomeric repeat sequence is 5'-TTAGGG-3', consisting of a double-stranded

region ranging between 5-50 kb and a short single-stranded G-rich 3'-overhang with a length of 100-300 nucleotides.[137,138] Since the G-rich strand extends beyond its complementary C-rich strand, it is proposed that it forms a T-loop (telomeric loop), wherein the 3'-overhang invades the double-stranded telomeric region, causing strand displacement of the G-strand bound to the C-rich strand. This forms a second loop called the D-loop (displacement loop, **Figure 1.10B**). Thus, the T-loop caps the telomeric ends and protects telomeres from the DNA damage response.[139] Since the repetitive G-rich overhang is free of its complementary C-rich strand, these overhangs could fold into G4 structures (discussed in depth in Section 1.6).[140]

Telomeres are considered dysfunctional either when they reach a critically-short length or when any of the shelterin complex components is delocalized from the telomeres. Telomerase, the ribonucleoprotein responsible for telomere maintenance, extends telomeres to counteract telomere shortening and avoid end fusion of chromosomes, which leads to genomic instability. Telomerase is overexpressed in 85% of cancer cells, but is absent in the remaining 15% where telomere elongation occurs *via* the alternative lengthening of telomeres (ALT) mechanism based on the recombination between telomeres.[141] In 2009, Elizabeth H. Blackburn, Carol W. Greider, and Jack W. Szostak shared the Nobel Prize in Physiology or Medicine for discovering the enzyme telomerase and how chromosomes are protected by telomeres.[142]

1.5.2 Telomerase and Telomere Maintenance

Telomerase consists of a Telomerase Reverse Transcriptase protein (hTERT)[143] and an RNA component (hTR).[144] Absence of telomerase expression in most somatic cells (with the exception of stem cells, skin keratinocytes, germ cells, *etc*.) is considered a tumor suppressing mechanism, while telomerase is overexpressed in approximately 85% of human cancer cells.[145] Therefore, telomerase is considered an important cancer biomarker and a potential therapeutic target. The mechanism of telomere extension by telomerase is outlined in **Figure 1.11**. Telomerase binds to the tip of the G-rich 3'-overhang, using it as a primer. hTERT catalyzes reverse transcription, using hTR as the template, and leads to nucleotide addition in the 5' to 3' direction. When the end of the template is reached, telomere extension continues, with telomerase (*i.e.* the RNA template) translocating by 6 nucleotides. In order to fill in the gaps in the complementary C-rich strand, a primer is synthesized, and then DNA polymerase α, which

possesses DNA primase as one of its subunits, fills in the gaps. The primer is then removed and ligase seals the gaps between the synthesized fragments.[135]

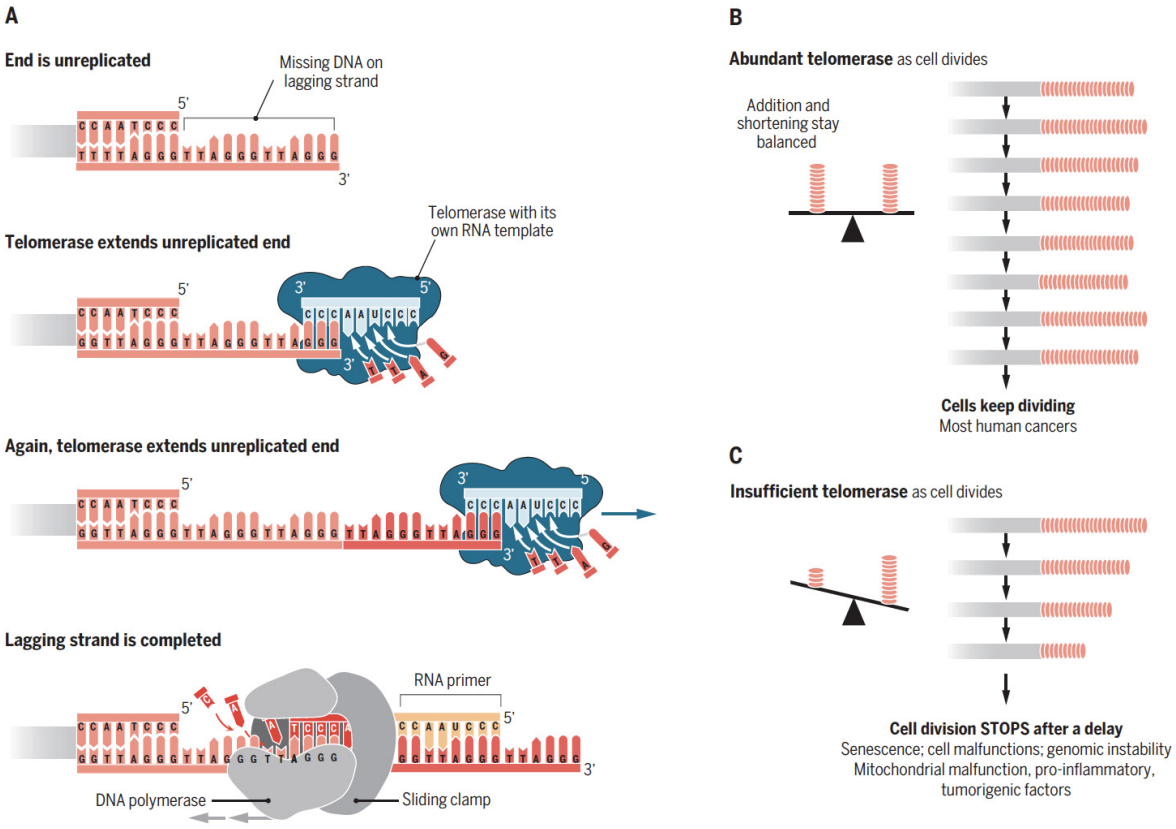


Figure 1.11: A) Mechanism of telomere elongation by telomerase. B) In the presence of telomerase, telomere shortening during cell division is balanced with telomere-elongation by telomerase. C) Lack of telomerase leads to critically short telomeres and eventually cell death. Figure adapted from reference [135a].

The ability of telomerase to add several repeats to the same telomeric strand is known as "repeat addition processivity" (RAP).[146] RAP is a unique feature of telomerase that is taken advantage of in telomerase activity assays that measure telomerase activity *in vitro* by monitoring the extension products. Several assays have been developed thus far and classified as PCR-based or direct (without amplification) assays. TRAP (Telomere Repeat Amplification Protocol) is a widely used technique with great sensitivity but suffers some limitations.[147,148] Most importantly, TRAP was proven disadvantageous for determining the effect of G-quadruplex ligands on telomerase inhibition. These ligands might lead to false positives due to their interference with the PCR amplification step; thus leading to inhibition of telomeric DNA amplification rather than inhibition of telomere extension.[149] In order to circumvent this

limitations, direct telomerase activity assays were developed. These assays utilize purified telomerase, which would lead to the removal of polymerase inhibitors and allows for the preparation of more concentrated samples of the enzyme, thereby enhancing the specificity of the assays.[146] Several approaches have been developed in order to inhibit telomerase by targeting its RNA template or catalytic components such as oligonucleotide inhibitors complementary to the RNA template, nucleoside and non-nucleoside reverse transcriptase inhibitors, and G-quadruplex ligands.[141]

1.6 G-Quadruplexes: Structure, Function, and Biological Relevance

G-quadruplex structures have gained considerable research interest due to their thermodynamic stability near physiological conditions, unlike other non-canonical secondary structures. They have been utilized in several areas of research such as chemistry, nanotechnology (self-assembly),[150] biotechnology (DNAzymes),[151,152] sensing,[153] aptamers,[117] and as potential anti-cancer drug targets.[154]

1.6.1 G-tetrads: The Building Blocks of G4 Structures

The G-tetrad, also known as G-quartet, was initially discovered in 1962 from the aggregation of 5'-monophosphate guanosines into guanosine gels.[155] A G-quartet is a square planar arrangement of four guanine bases connected through Hoogsteen and Watson-Crick hydrogen bonds between N1–O6 and N2–N7 (**Figure 1.12A**). In the tetrad, every guanine is rotated around 90° with respect to its neighbor, thereby positioning the O6 atoms in the center of the tetrad. The self-stacking of the planar G-tetrads leads to the formation of G4 structures and extended 3D architectures stabilized by a central spine of cations.[156] G4 structures generally exhibit a right-handed helical nature with a rise of 3.13–3.3 Å, a 30° right-handed twist, and a 25 Å diameter (**Figures 1.12B**).[157] However, in 2015, the Phan group discovered the first left-handed intramolecular G4 structure of the $d(T(GGT)_4TG(TGG)_3TGTT)$ oligonucleotide *via* X-ray analysis and NMR spectroscopy.[158]

The core of the G4 structure can accommodate a variety of mono and divalent cations of different radii coordinated by the carbonyl groups of the guanine bases.[159] The nature of the cation poses a significant influence on the stability and the folding topology of G4 structures. This effect was first discovered by Sen and Gilbert in 1990.[160] The stability of G4 structures has

been assessed in the presence of several cations, and the ranking based on the stabilization of G4 structures is: $Sr^{2+} > Ba^{2+} > K^+ > Ca^{2+} > Na^+$, $NH_4^+ > Rb^+ > Mg^{2+} > Li^+ \geq Cs^+$.[156] The radii of the divalent metal ions, Sr^{2+} and Ba^{2+}, allows them to fit in the central core of the quadruplexes leading to significant stabilization.[161] Among monovalent cations, Na^+ and K^+ have gained considerable attention due to their high intracellular concentrations (~100-150 mM and 10 mM, respectively). K^+, the most stabilizing monovalent cation, is usually flanked by two G-tetrads and coordinates with eight guanine carbonyl groups.[162] On the other hand, Na^+ is less stabilizing and is usually localized in the plane of the G-tetrad coordinated with four carbonyl groups.[163] The thermodynamic stability of G4 structures is directly proportional to the concentration of monovalent ions.[164]

1.6.2 G-Quadruplexes: A Single Building Block Leading to a High Degree of Structural Polymorphism

Even though G4 structures consist of just one building block, guanosine, they have a high degree of structural polymorphism compared to B-DNA, which consists of 4 building blocks (A, T, G, and C). This polymorphism is due to several factors, including the number and orientation of the strands forming the G4 structures, the nature of the cation, the length of the G-tracts, the length and composition of the loops connecting the G-tracts, and the glycosidic bond orientation (*syn* or *anti*).[165] G4s form in DNA and RNA strands *via* intramolecular or intermolecular interactions.

In a parallel G4 topology, all the G-tracts are aligned in the same direction (*i.e.* 4 parallel + 0 antiparallel) and the loops connecting neighboring G-tracts on opposite surfaces are called double-chain-reversal or propeller loops (**Figure 1.12C**).[166,167] When three out of four G-tracts are parallel (*i.e.* 3 parallel +1 antiparallel), the G4 structure is known as mixed or hybrid (**Figure 1.12D**) and possesses propeller, diagonal, and lateral loops. Diagonal loops connect antiparallel strands from opposite edges of the G4 structure on the same surface, while lateral loops link neighboring antiparallel G-tracts.[157] An antiparallel G4 structure consists of two parallel and two antiparallel strands. Two antiparallel topologies have been characterized: the basket possessing lateral-diagonal-lateral loops (**Figure 1.12E**), and the chair possessing three lateral loops (**Figure 1.12F**). The type of loops connecting the G-tracts depends on the number and nature of nucleotides in the loop. Propeller loops in parallel G4 structures are usually short in length (1-2

nucleotides), while lateral and diagonal loops are longer. Therefore, the nature of the loops plays a major role in determining the topology and stability of G4s.

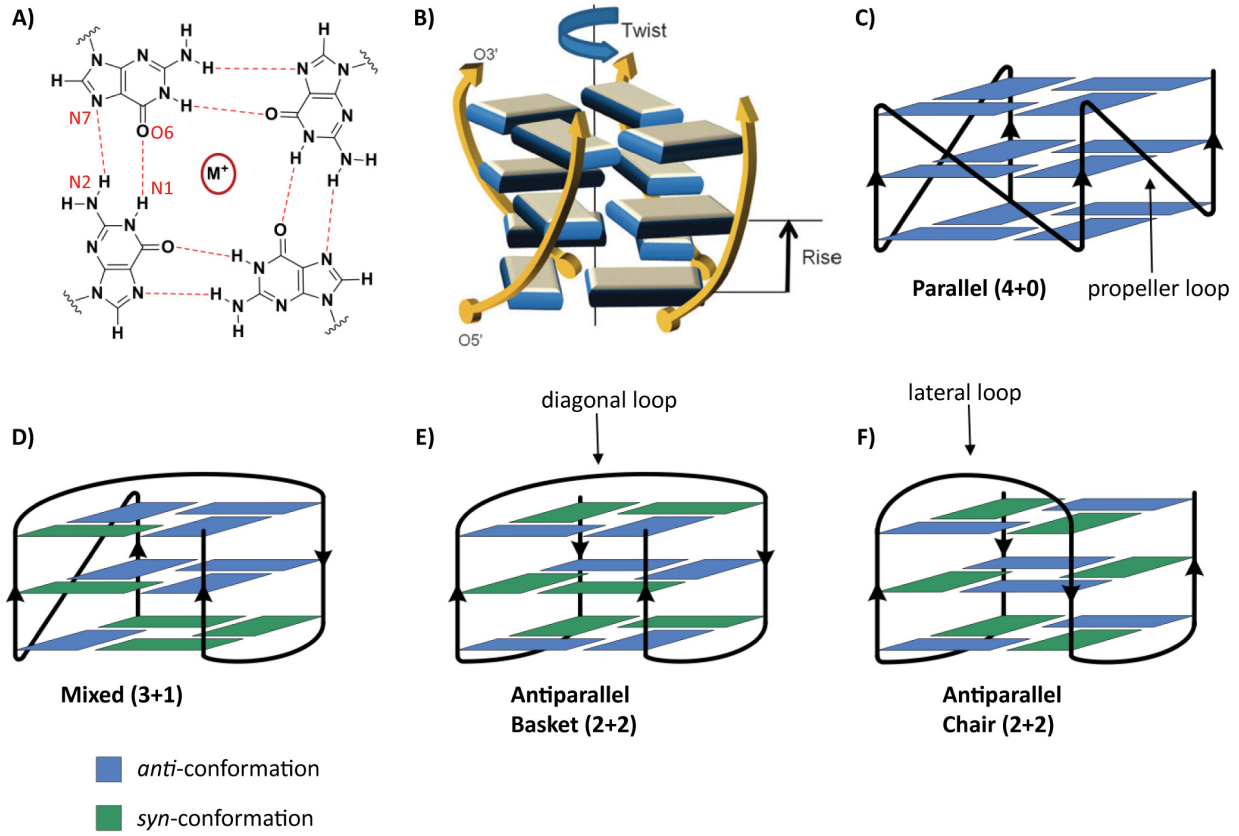


Figure 1.12: A) Structure of a planar G-tetrad. B) Helical parameters (twist, rise, and backbone polarity) for a tetramolecular G4 structure, adapted with permission from reference [157] (Royal Society of Chemistry, 2011). C-F) Different G4 topologies and the loops connecting the G-tetrads in each G4 structure.

Another factor that plays a major role in G4 folding is the glycosidic bond angles connecting N9 of the guanine base in the G-tetrad to the C1' of the ribose sugar. The torsion angle can either be *syn* (0 to 90^{o}) or *anti* (-120 to 180^{o}).[157] In a parallel G4, all glycosidic angles are *anti*, while in antiparallel and hybrid G4 structures, the bases are a mix of *syn* and *anti*.[168]

In addition to the *syn*/*anti* conformation of guanines and the type of loops, the nature and concentration of the cation in the medium has a significant impact on the folding of G4 structures. The four-repeat telomeric sequence was found to adopt different topologies depending on whether Na^+ or K^+ was utilized. In 1993, the Patel group was the first to characterize the solution NMR structure of d[AGGG(TTAGGG)$_3$] in Na^+.[169] The G4 folding topology was found

to be intramolecular, comprising three stacked G-tetrads with *anti-anti-syn-syn* glycosidic bonds and three TTA loops (lateral, diagonal, and lateral). In 2002, Neidle and Parkinson obtained the crystal structure of the same sequence in K^+. The crystallized G4 structures exhibited a parallel topology; therefore, all loops were double-chain-reversal (propeller) loops and all guanines adopted an *anti* glycosidic conformation.[170] On the contrary, several laboratories reported that in the presence of K^+ in solution, the human telomeric repeat exists as a mixture of conformations.[171-174]

Unlike telomeric sequences that are highly repetitive, G4-forming sequences within promoter regions possess G-tracts with unequal numbers of guanines and diverse loop sequences. This leads to an exchange between several G4 structures with different G4 cores due to a shift in the alignment of the G-tracts. This exchange is called G-register (GR) exchange dynamics and the different topologies are known as GR isomers. For instance, the 18mer c-MYC Pu18 sequence, 5'-AGGGTGGGGAGGGTGGGG-3', was found to exchange between four different GR isomers.[175] Another interesting type of GR exchange may result from the replacement of an entire damaged G-tract with a "spare tire".[176] The Burrows group discovered that when a fifth G-tract is present a few nucleotides away from an oxidatively-damaged G4 structure in oncogene promoter regions, it can act as a "spare tire" and entirely substitute the damaged G-tract. Damaged G-tracts are due to inflammation and oxidative stress that lead to the oxidation of the guanines in G4 structures formed in several oncogenes. When the damaged guanines, such as 8-oxo-7,8-dihydroguanine (OG) and 5-guanidinohydantoin (Gh), are present in the four G-tracts of a G4 fold, they are not removed by base excision repair (BER) DNA glycosylases.[177] The replaced G-run (*i.e.* the spare tire) is now a substrate of BER processes, thus suggesting a role of G4 structures as sensors of oxidative stress and regulators of gene expression.

1.6.3 Trapping G-Quadruplex Geometries Using Chemically-Modified Nucleotides

The polymorphism observed in G4-forming sequences complicates the structural determination of specific G4 topologies by NMR spectroscopy. Since the orientation of the guanine nucleobases directly influences the G4 structure, researchers have taken advantage of chemically-modified nucleotides that favor either the *syn* or the *anti* base conformation to engineer G4 folding topology and slow down topological exchange.[178-182]

While DNA telomeric G4s adopt a plethora of polymorphs, the human telomeric RNA G4 structures always adopt a parallel topology due to the presence of the 2'-hydroxyl groups.[183] The 2'-hydroxyl group in RNA stabilizes the C3'-*endo* sugar pucker and locks the glycosidic bond in the *anti* conformation, thereby favoring the parallel folding topology.[157] Moreover, these hydroxyl groups participate in intramolecular hydrogen bonding interactions (*i.e.* between O4' of the ribose sugars and with N2 groups of the central guanines) instead of interacting with water molecules, resulting in RNA quadruplexes that are more thermodynamically stable than their DNA counterparts.[184] Substitution of the 2'-H with 2'-O-Me rather than 2'-OH destabilizes the G4 structures formed from a thrombin-binding aptamer and from an 18-mer telomeric repeat.[180]

Our group and others have studied the effects of 2'F-ANA and 2'F-RNA on G4 structure and stability.[181,182,185,186] 2'F-araG, araG, 2'F-riboG, and riboG modifications were introduced in the two-repeat telomeric sequence, d(TAGGGTTAGGGT), which normally adopts parallel propeller and antiparallel conformations in solution.[171] Since guanine at position nine adopts a *syn*-glycosidic bond conformation,[168] it was hypothesized that a modified nucleotide with *anti* conformation could shift the equilibrium of the G4 topologies towards the parallel propeller fold. 2'F-araG favors the *anti* glycosidic bond orientation to avoid the steric clash between the guanine base and the fluorine in the arabino position (**Figure 1.13A**).[181]

It was observed that 2'F-araG leads to the most stable intermolecular parallel structure (ΔT_m = +12 °C) due to electrostatic (F−CH---O4') and pseudo-hydrogen-bonding (F---H8) interactions (**Figure 1.13)**, while 2'F-riboG substitution is extremely destabilizing due to unfavorable electrostatic repulsions between the fluorine at the 2' position and the oxygen atoms in the phosphate backbone. Therefore, a single dG to 2'F-araG substitution is sufficient to selectively stabilize the parallel topology over competing conformations. With respect to scope, 2'-fluoro modifications do not simply trigger a conformational switch; they can also be utilized as ^{19}F-NMR probes to aid in structural determination.[182,185] LNA modifications favoring the *anti* glycosidic conformation of the guanine base have also been introduced in G4 structures. Similar to the 2'F-araG scenario, LNA modifications shift the structural equilibrium towards the parallel conformation when introduced into positions adopting *syn* glycosidic bonds.[186] Moreover, LNA modifications have been shown to enhance the biological activity of G4 structures.[186]

Several natural and non-natural nucleobase modifications have also been investigated.[179] Bulky substituents at the C8 position of guanosine, such as 8-bromoguanine, destabilize the *anti* glycosidic bond and shift the equilibrium conformation to favor the *syn* orientation by at least 1-2 kcal/mol.[187,188] Therefore, these modifications can only be utilized in structures possessing *syn* glycosidic bonds.

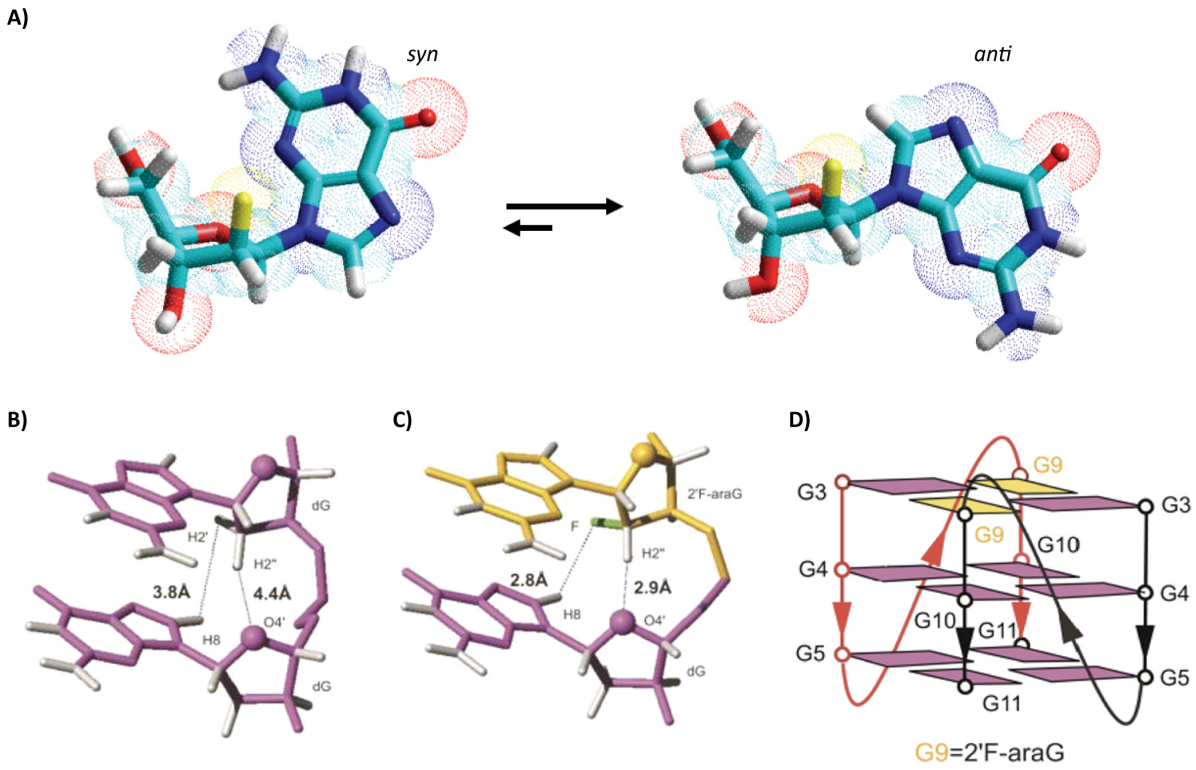


Figure 1.13: A) The guanine in 2'F-araG adopts an *anti* glycosidic conformation. Figure adapted from reference [181] (Nucleic Acids Research, open access). B) Distances in the crystallographic structure (PDB code: 1K8P) between H2'−H8 and H2"−O4'. C) Distances between F−H8 and H2"−O4' in the 2'F-araG modified structure. Guanines in dG are shown in magenta and yellow in 2'F-araG, and fluorines are shown in light green. D) Scheme of the parallel propeller G4 with all guanines in the *anti* conformation. 2'F-araG nucleotides are shown in yellow and dG residues in magenta. B-D) Adapted with permission from reference [185] (American Chemical Society, 2013).

In general, nucleobase modifications in the G-tetrad are destabilizing, especially if they interfere with the hydrogen-bonding network across the tetrad.[189] Nucleobases resulting from oxidative stress, such as 8-oxo-7,8-dihydroguanine, do not only induce structural switching of G4 structures in oncongene promoters by utilizing their "spare tire",[176] but also lead to the up- or down-regulation of transcription depending on whether they are located on the template or on the coding strand of the promoter.[190]

1.6.4 Location of G-Quadruplex-Forming Sequences Throughout the Genome

Given the growing interest in investigating G4 structures, several algorithms have been developed in order to locate G4-forming sequences across the genome. The initial prediction tool utilized was the "Quadparser" algorithm, $G_{\geq 3}N_{1-7}G_{\geq 3}N_{1-7}G_{\geq 3}N_{1-7}G_{\geq 3}$, where N = A, T(U), G, C.[191,192] This algorithm determined around 300,000 sequence motifs across the genome with the tendency to fold into G4 structures.[193,194] Subsequent studies discovered numerous motifs that fold into G4 structures such as G4s in minisatellite DNA possessing a 9-nt middle loop,[195] the parallel topology of the BCL-2 promoter gene containing a 13-nt loop,[196] and G4 structures with bulges of varying length across the G4 core;[197] however, these structures deviate from the "Quadparser" algorithm. False positives were also discovered following Quadparser screening, such as G-rich sequences in the 5'-UTR (untranslated regions) in mRNA.[198] To circumvent these limitations, the development of a better prediction tool was necessary to take into account a wider range of criteria. In 2016, the Mergny group developed the G4Hunter algorithm, which accounts for the G-richness, G-skewness, and G/C asymmetry on the complementary strand of a studied sequence.[199] Another method utilized to estimate the number of G4 structures was based on G4-dependent DNA polymerase stalling combined with next-generation sequencing (G4-seq) of single-stranded DNA.[200] In this method, G-rich strands were sequenced under conditions that destabilize G4 structures and conditions that favor G4 folding. G4-dependent polymerase stalling and a significant decrease in sequencing data were observed in the G4 rich constructs. Based on the G4-seq method, more than 700,000 G4 structures have been detected in the human genome.[200]

Studies show that putative G4-forming sequences are highly populated in oncogene promoter regions (c-MYC,[201,202] BCL-2,[196] c-Kit,[203,204] VEGF,[205] RET),[206] in transcription factors Nrf2,[207] SRC,[208] in the promoter of the human telomerase reverse transcriptase gene (hTERT),[209] in 5'-UTR mRNA, and in G-rich telomeric repeat RNAs. G4-forming sequences also exist in the genomes of numerous viruses.[191,210,211] A breakthrough discovery in the G4 field was the development of G4-specific antibodies by the Balasubramanian group which confirmed the presence of G4 structures in human chromosomes.[212] These antibodies did not only target a specific G4 topology, but were also capable of differentiating parallel G4 topologies with 100-fold specificity, even with only minor differences in the loop structure. Furthermore, quantitative visualization of DNA and RNA G4 structures in cells was achieved with BG4 antibodies in

immunofluorescence assays (**Figure 1.14**).[213,214] In these studies, BG4 antibodies were scattered in several chromosome loci (**Figure 1.14**, i–iii) and at telomeric ends (**Figure 1.14**, iv and v). Interestingly, G4 structures were observed in sister chromatids of newly replicated DNA (**Figure 1.14**v), highlighting the dependence of G4 formation on the stage of cell division with maximum G4 levels obtained during the S-phase.

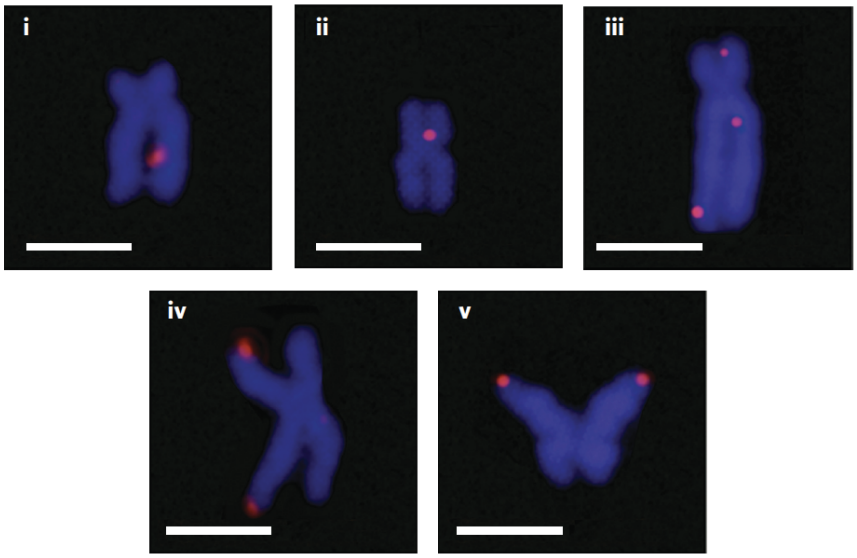


Figure 1.14: Immunofluorescence for BG4 antibodies on metaphase chromosomes isolated from Hela cervical cancer cells leading to the observation of G4 structures in chromosomes. Scale bars, 2.5 mm. Adapted with permission from reference [213] (Nature Publishing Group, 2013).

1.6.5 Biological Significance of G-Quadruplex Structures

X-ray and NMR structural determination of numerous G4 folding topologies combined with the imaging and mapping of these structures throughout the genome have provided significant insight with respect to their role in several biological processes including transcription, replication, and telomere maintenance and elongation. The formation and stability of G4 structures in oncogene promoter regions gives them an important role in regulating gene expression. For instance, stabilizing G4 structures with ligands or synthetic oligonucleotides in the proto-oncogenes c-MYC,[215] KRAS,[216] and zebrafish embryos leads to decreased mRNA levels accompanied by phenotypic changes,[217] while destabilizing G4 structures in promoter regions leads to an increase in gene expression.[218,219] Introducing mutations in the G4-resolving helicases alters the expression of the G4-rich promoter regions, thus confirming the influence of G4 structures on transcription.[220,221] The presence of stable G4 structures, due to the disruption or

absence of helicases, inhibits the progression of DNA polymerases, leading to genomic instability.

G4 structures also have important consequences in telomere biology. Folding of the G-rich 3'-overhang at the chromosome ends into G4 structures can prevent the hybridization of the RNA template of telomerase to the 3'-overhang, thus blocking the extension of telomeric DNA.[222] G4s are also thought to be involved in inhibiting telomere recombination, thereby impeding the alternative lengthening of telomeres.[223] Moreover, G4 structures might hinder the association of the shelterin complex proteins to the 3'-end and lead to telomere dysfunction. Stabilization of telomeric G4 structures *via* small-molecule ligands has been considered as an interesting strategy for anticancer therapy.[224]

Early studies reported that G4 structures prevent the extension of telomeres by telomerase but did not take into account that G4 conformation could influence its interaction with telomerase.[225] In 2006, the Bryan group reported that the parallel-stranded intermolecular G4 structures from the ciliated protozoa *Tetrahymena thermophila* and *Euplotes aediculatus* are elongated by the ciliate telomerase, while the intramolecular antiparallel conformations are not substrates of telomerase.[226] In 2015 the same group showed that the ability of telomerase to extend parallel G4 structures is not a ciliate-specific property.[227] They demonstrated that human telomerase is able to partially unwind and extend parallel intermolecular G4 structures (**Figure 1.15**). In sum, these results show that not all G4 structures lead to inhibition of telomerase, as was originally thought.

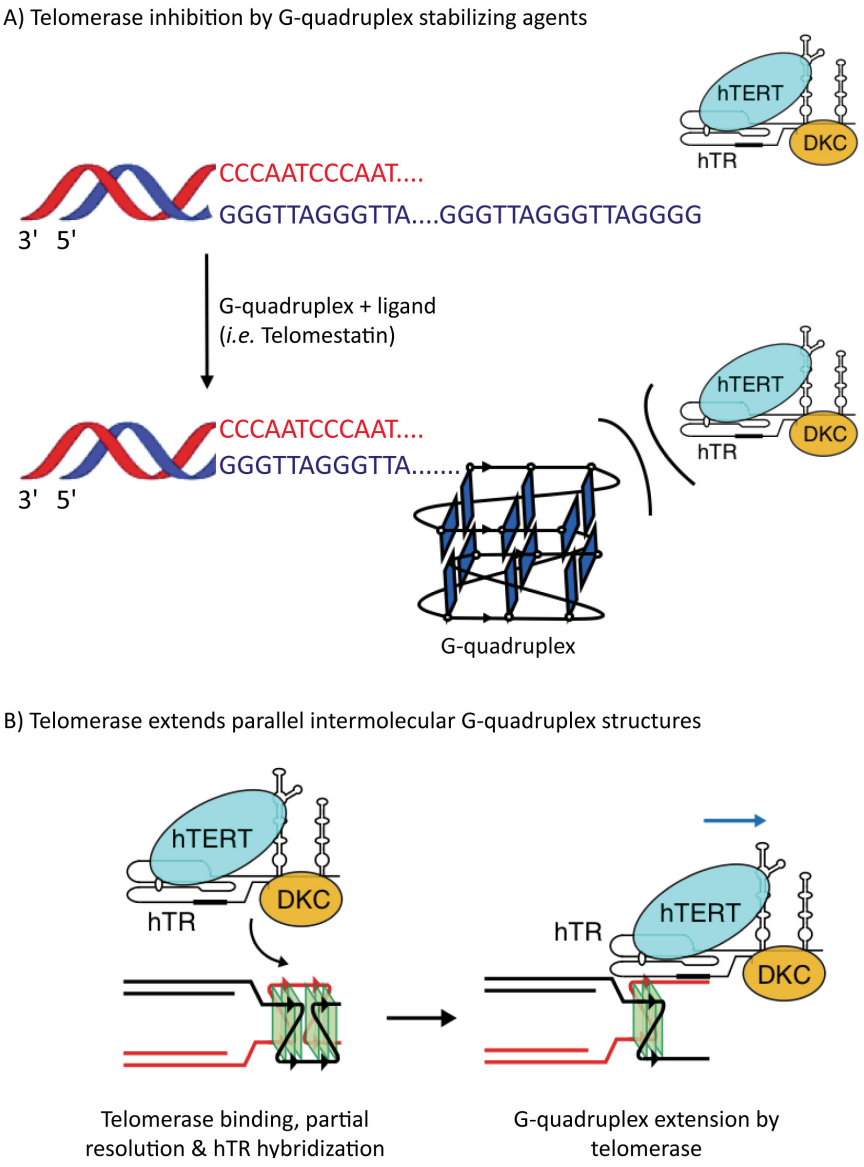


Figure 1.15: A) Inhibition of telomerase activity due to G4 formation. B) Proposed mechanism for the elongation of parallel intermolecular G4 structures by telomerase, adapted from reference [227] (Nature Communication, open access).

1.6.6 Ligands and Proteins that Recognize and Bind to G-Quadruplex Structures

Structural polymorphism observed for G4 structures can be a bottleneck for the design of selective ligands. Small-molecule ligands targeting G4 structures should possess certain features. For instance, they should exhibit a large aromatic core in order to enhance the π-stacking interactions with the surface of the G-quartet, a positive charge to counteract the negatively charged backbone, and side chain functionality to provide additional interactions with the G4 grooves, loops, or bases.[228] These ligands can either bind to pre-formed G4 structures or initiate

the folding of G-rich sequences into G4 structures (molecular chaperone).[208] The ability of ligands to discriminate between various G4 topologies is extremely important for their application as therapeutic agents.[229] BRACO19,[230] RHSP4,[231] and telomestatin[232] are examples of three ligands that bind to telomeric DNA. For instance, telomestatin, a natural product from *S. Anulatus*, was found to slow the growth of glioblastoma since it alters the binding of shelterin protein TRF2 to telomeric DNA which normally results in glioma stem cells.[189] Additionally, the discovery of several proteins that specifically bind to DNA[141] and RNA G4s is extremely significant since it further supports the biological relevance of these secondary structures.[233,234] Three different types of G4-binding proteins have been reported so far: telomere-binding proteins (*i.e.* the shelterin complex);[233,235] nucleases and helicases responsible for G4-unwinding;[236,237] and proteins that stabilize G4s or act as molecular chaperones such as nucleolin and nucleophosmin.[238,239]

The high thermodynamic stability of G4 structures at physiological conditions followed by the *in vivo* visualization of these structures in DNA and RNA led to a tremendous leap in the G4 field. As a result of Watson-Crick base pairing, every G4-forming sequence has a complementary C-rich strand with the tendency to form i-motif structures, also known as i-DNA or i-tetraplexes.

1.7 Nucleic Acid i-Motifs: Structure, Stability, and Targeting Ligands

The i-motif field has grown at a slower pace compared to the G4 field because i-motif structures are stabilized by hemi-protonated base pairs under acidic conditions (**Figure 1.16A**). Despite the identification of hemi-protonated cytosine-cytosine$^+$ base-pairing in 1962 in crystals of cytidine-5-acetic acid and the first i-motif structure in 1993,[240,241] it was not until recent years that the i-motif field started to develop as evidence emerged for their biological relevance in transcription, replication, telomere functions, and telomerase inhibition. The pH-dependence of i-motif structures furthered their use as pH switches in nanotechnological applications,[242] supramolecular structures, logic gates,[243] molecular beacons,[244] and pH sensors in living cells.[245]

1.7.1 i-Motif Structure and Topology

The first i-motif structure was characterized by Gehring *et al.* for the hexamer sequence d(TCCCCC) forming an intercalated quadruple-helical tetramolecular structure under acidic conditions (**Figure 1.16B**).[241] The structure consists of two parallel-stranded duplexes intercalated in an antiparallel orientation and held together by hemi-protonated cytosine-cytosine$^+$ base-pairing.[241,246] Therefore, the building block of an i-motif structure consists of a base pair between a neutral cytosine and a cytosine protonated at position N3 leading to three hydrogen bonds (**Figure 1.16A**). The base-pairing energy (BPE) for the C·C$^+$ base pair is 169.7 kJ/mol, compared to the BPEs of G·C (96.6 kJ/mol) and neutral C·C (68.0 kJ/mol).[247] The high BPE indicates that the hemiprotonated C·C$^+$ base pairs have stronger hydrogen-bonding interactions compared to their neutral counterparts and the canonical Watson-Crick G·C base pairs. Similar to G4 structures, i-motifs may fold in an intermolecular fashion from the association of two (dimers) or four (tetramers) strands, or form an intramolecular structure due to the spatial arrangement of C·C$^+$ base pairs from different C-tracts within the same strand.

The hydrogen bonding in a hemiprotonated C·C$^+$ (N3···H···N3) has been described as a double-well potential where the proton delocalizes/oscillates between the two wells.[248] Leroy *et al*. postulated the proton-transfer rate to be $8x10^4$ s^{-1}.[249] The NMR structural study of an intramolecular telomeric i-motif (PDB code: 1ELN) showed that the C·C$^+$ base pairs are planar and the N3–N3 distance is around 2.6 – 2.8 Å.[250] NMR spectroscopy and quantum chemical investigation performed by Lieblein *et al.* suggested that the N3···H$^+$···N3 bonds possess an asymmetric double-well potential and that the proton in one C·C$^+$ base pair tends to adopt a position that leads to the largest distance with respect to the proton of the consecutive C·C$^+$ base pair.[248]

Unlike other secondary structures such as B-DNA, G-quadruplexes, or hairpins that are held together *via* π stacking interactions between sequential nucleobases, i-motif structures lack stacking interactions due the intercalation between the consecutive base pairs. As shown in **Figure 1.16C**, no overlap exists between the six-membered aromatic pyrimidine nucleobases; however, C·C$^+$ base pairs involve the stacking of exocyclic carbonyls and amino groups in an antiparallel fashion.

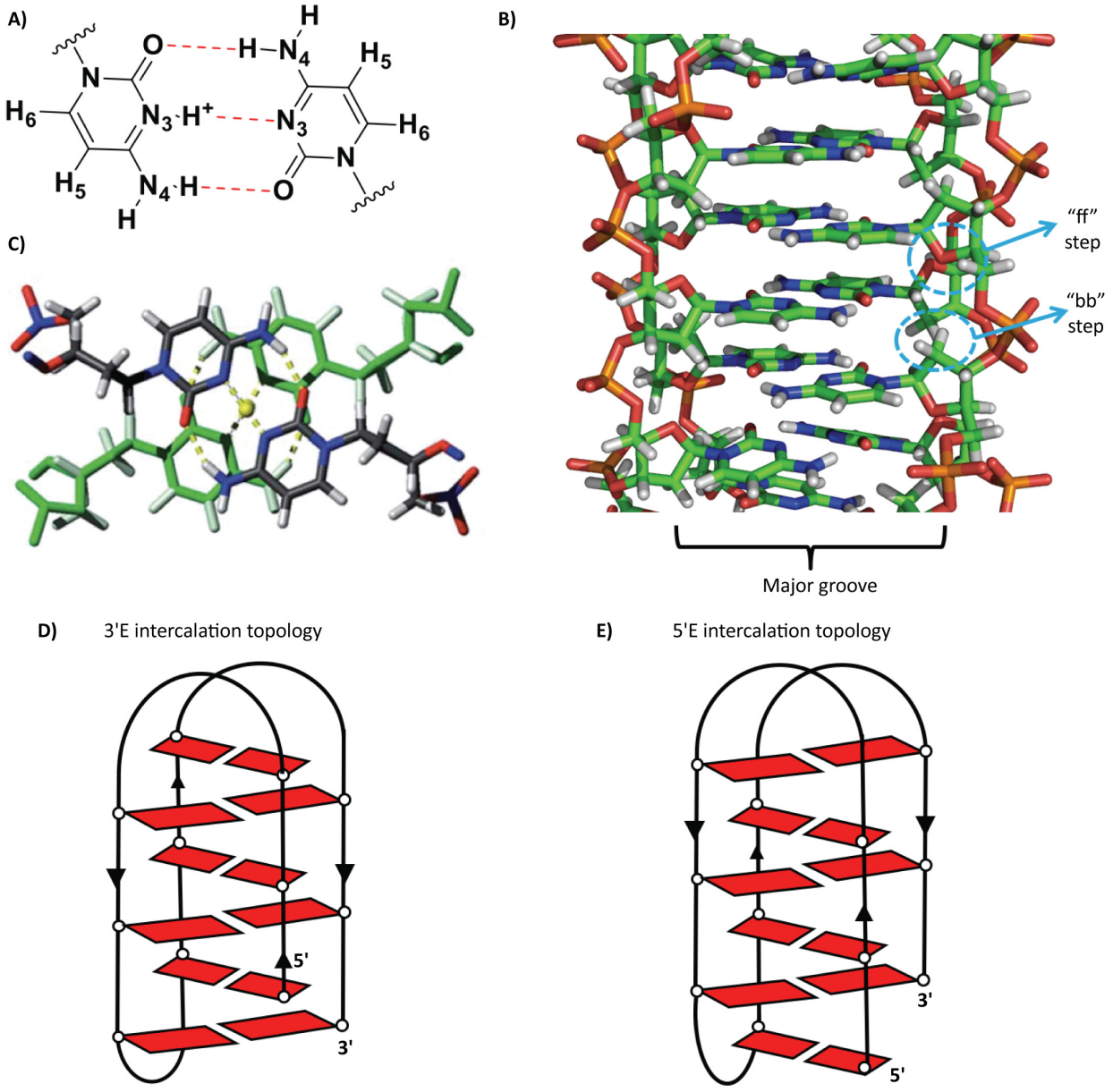


Figure 1.16: A) C·C$^+$ base pairs. B) 3D structure of a tetramolecular i-motif (PDB 1YBR), face-to-face and back-to-back steps are highlighted. C) Stacking between intercalating C·C$^+$ base pairs. Adapted with permission from reference [251] (Royal Society of Chemistry, 2014). D) Scheme representing the 3'E intercalation topology of an intramolecular i-motif structure. E) Scheme representing the 5'E intercalation topology of an intramolecular i-motif structure.

The structure consists of two wide major grooves and two narrow minor grooves. The intercalation of base pairs from two parallel duplexes leads to a shorter base-pairing distance (3.1 Å) and a smaller right-handed helical twist (12-16°) compared to A- and B-DNA, giving rise to close intermolecular contacts between the sugar phosphate backbones.[252] The minor groove consists of alternating face-to-face (ff) and back-to-back (bb) steps. In the ff steps, the O4' ring

oxygens face one another (5'-side); however, in the bb steps, the C3'-C4' edge of the sugar moieties are oriented close to each other (3'-side, **Figure 1.16B**). This unusual folding architecture is destabilized by two main factors: 1) the lack of base stacking interactions and 2) the close inter-strand phosphorus-phosphorus distances (5.9 Å) along the narrow groove.[253] These destabilizing factors are compensated by an intermolecular hydrogen-bonding network between the deoxyribose sugar moieties along the minor groove.[252] A C-H···O hydrogen-bonding network is observed between the lone pair of the O4' oxygen atom of a deoxyribose from one strand and H1' of a deoxyribose from the antiparallel strand along the narrow groove and vice versa. Moreover, studies have shown that a lone electron pair of one O4' can be shared between H1' and H4' hydrogens.[252] These very close contacts between the sugar-phosphate backbones from the antiparallel duplexes along the narrow groove lead to strong NOE signals between inter-strand sugar protons H1'-H1' and H1'-H4'.

Additional stabilizing interactions were observed in the crystal structures of d(CCCT) and d(CCCAAT), such as the intra-nucleoside hydrogen bond between the β lone pair of O4' and H6 in pyrimidine nucleobases (C6-H6···O4'). In addition to the C-H···O hydrogen-bonding network, the structure is stabilized by dipole-dipole and ion-dipole interactions between intercalated $C{\cdot}C^+$ base pairs.[241] All the mentioned interactions contribute significantly to the stability of i-motif structures.

Due to the spatial arrangement of $C{\cdot}C^+$ base pairs, d(TCCCCC) interconverts between two tetramolecular i-motif structures with different intercalation topologies known as 3'E and 5'E.[254] When the outmost $C{\cdot}C^+$ base pair is at the 3'-end, the structure is known as 3'E, while in the 5'E topology, the terminal $C{\cdot}C^+$ base pair is at the 5'-end.[127] Similarly, intramolecular i-motif structures form 3'E and 5'E intercalation topologies (**Figure 1.16D** and **E**). NMR studies on the intramolecular human telomeric repeat $d(CCCTAA)_n$, indicated that the 3'E configuration is kinetically favored while the 5'E topology is thermodynamically more stable due to an extra T-T base pair between the first and third loop.[255]

1.7.2 Factors Affecting the Stability of i-Motif Structures

Detailed investigation of the factors that influence i-motif folding and stability is important in understanding how i-motifs regulate gene expression and how to implement these

structures efficiently in nanotechnological applications. Like other nucleic acid structures, i-motif stability depends on several factors, including sequence nature and length, C-tract length, loop topology, temperature, ionic strength, ligands, and pH. Initially, it was thought that i-motif structures could only fold at acidic pH values, due to the requirement of hemiprotonated base pairs; however, several recent studies have shown that i-motif structures might form at neutral pH depending on the sequence and environmental conditions. For example, i-motif structures were observed at neutral pH and low temperatures under molecular crowding conditions,[256,257] under negative superhelicity,[201] and in the presence of silver cations.[258] Moreover, as shown in this thesis work, 2'F-araC chemical modifications stabilize inter- and intramolecular i-motif structures at neutral conditions and high temperatures.

1.7.2.1 C-tract Length and Genomic Analysis

In general, under the same experimental conditions, the i-motif structure possessing a higher number of $C{\cdot}C^+$ base pairs would be more stable.[259] Very recently, Zoe Waller's group and Cynthia Burrows' group investigated the effect of C-tract length on the folding of i-motif structures under physiological conditions.[260,261] Waller's group studied several sequences with varying C-tract lengths. Their results suggest that, at pH 7.4, the thermal stability increases as the number of cytosines per tract increases from three ($C_3(T_3C_3)_3$; T_m = 7 °C and pH_T = 6.7) to five $C_5(T_3C_7)_5$; T_m = 26.2 °C and pH_T = 7.2). Increasing the tract length beyond five enhances the stability; however, it also leads to two melting transitions and to increased hysteresis between the folding and unfolding transitions. Sequences with ten cytosines per tract $C_{10}(T_3C_{10})_3$ exhibit a pH_T of 7.3 and therefore still form i-motif structures at neutral pH. In sum, the study suggests that a minimum of five cytosines is required to achieve stability at neutral conditions.

Based on these findings, the Waller group utilized Quadparser to determine the potential prevalence of such i-motif-forming sequences within the human genome. They developed a folding rule, $C_5(N_{1\text{-}19}C_5)_3$, with four C-tracts of five cytosines and a loop length ranging between 1 to 19 nucleobases.[260] This rule led to the identification of 5125 sequences across the genome with the potential to fold into i-motif structures. Out of the 5125 i-motif structures, 637 (*i.e.* ~12.4%), were located in gene promoter regions. Thirty genes were found to have several i-motif-forming sequences within their promoter regions. Through careful examination of the gene ontology codes of 100 genes, it was found that i-motif formation was concentrated in promoter

regions and regulatory genes. More specifically, i-motif formation was concentrated in "genes involved in skeletal system development, sequence specific DNA binding, DNA templated transcription and positive regulation of transcription from RNA polymerase II promoter" and not in "genes involved in the immune response, G-protein coupled receptor activity and olfactory receptor activity".[260]

On the other hand, the Burrows group was interested in investigating the formation and stability of i-motif structures from several dC_n strands (where n = 10 – 30).[261] Several sequences exhibited high pH_T and T_m values, suggesting their ability to form i-motif structures under physiological conditions. The highest stability was observed for sequences containing 15, 19, 23, and 27 cytosines. This led to the determination of a 4n − 1 "sweet spot" for i-motif folding in deoxycytidine homopolymers. By introducing T nucleotides to form loops of varying lengths, it was determined that the highest stability is achieved with an even number of $C{\cdot}C^+$ base pairs in the core and three loops consisting of a single nucleotide each. However, it is important to note that these results are not applicable for hetero-oligonucleotides since the presence of loop-to-loop interactions might contribute to the stability of the structures.[262]

Using bioinformatics analysis, the Burrows group investigated the presence of dC_n tracks across the human genome. These studies suggested the presence of 269 dC_{15} sequence motifs and 769 dC_n tracks with n between 15 to 81 nucleotides. Consistent with the results obtained from Waller's group, these sequences were mainly distributed in promoter regions, introns, and 5'- and 3'-UTRs. On the contrary, fewer dC_n strands were observed in the coding and intergenic regions. These two studies are not comprehensive, but they are an optimistic starting point in determining the number and occurrence of i-motif-forming sequences in the genome. They also highlight the fact that these sequences are not randomly located; instead, they are localized in the promoters of several genes and might have a great impact in the activation or suppression of genes.

1.7.2.2 Effects of Chemical Modifications on i-Motif Stability

1.7.2.2.1 Sugar Modifications

There is an increasing interest in stabilizing i-motif structures at a wider range of pH for biological and nanotechnological applications. Since i-motif stabilization is highly pH-dependent, chemical modifications have been introduced to modulate their stability. Several sugar modifications have been investigated, some of which are: RNA,[263,264] 2'-arabinose,[265] 2'F-RNA,[266] locked nucleic acids (LNA),[267] unlocked nucleic acids (UNA),[268] acyclic threoninol cytidine nucleic acids (*a*TNA),[269] and 2'F-araC (**Figure 1.17**).[270]

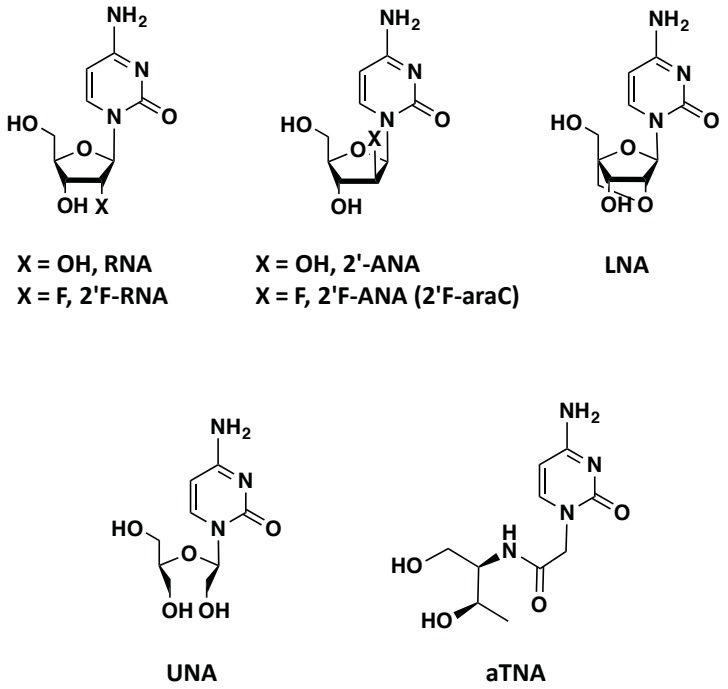


Figure 1.17: Sugar modifications introduced in i-motif structures.

In DNA i-motif structures, the glycosidic angles possess an anti-conformation and the deoxyribose sugars mainly adopt a C3'-*endo* (North) pucker. These properties led researchers to investigate the effect of RNA and RNA-like modifications locked in the North conformation on i-motif stability. Reports have shown that oligoribonucleotides form less stable i-motif structures compared to their oligodeoxynucleotide counterparts. Initial studies reported a ΔT_m of 29 °C between i-motif structures formed by an 18-mer DNA sequence, (d(CCCTCCCTTTTCCCTCCC), 54 °C) *versus* the corresponding 18-mer RNA sequence (r(CCCUCCCUUUUCCCUCCC), 25 °C).[264] On the contrary, the uracil substituted DNA

sequence exhibits a melting temperature of 56 °C. Therefore, the destabilization of RNA i-motif is not due to the lack of the methyl group on uracil but due to the presence of 2'-OH group. Collin and Gehring studied the effect of chimeric DNA/RNA tetrads on i-motif formation.[263] This study suggested that the destabilization observed due to an RNA residue is almost equal to the destabilization due to the loss of a $C{\cdot}C^+$ base pair. When two hydroxyl groups are juxtaposed in back-to-back step, the distance between them is 0.35 nm compared to 0.65 nm in face-to-face steps, leading to less stable i-motif structures (**Figure 1.18**).[263]

The r(UC_5) sequence led to the formation of two i-motif structures adopting different intercalation topologies. The intercalation topology of the major conformation possesses one less 2'-OH/2'-OH repulsive contact compared to six in the minor conformation. The free energy of the RNA i-motif per $C{\cdot}C^+$ base pair (-4 KJ/mol) is almost half that of the DNA i-motif.[271] Similarly, 2'-*O*-modifications completely perturb i-motif formation, even when only one insert is introduced. On the contrary, substituting the 2'-OH in RNA with its 2'-arabinose epimer leads to stable i-motif structures since the OH group is placed in the wide major groove.[265] The different stability observed between riboses and arabinoses confirms the critical significance of sugar–sugar contacts in the minor groove on the stabilization of i-motifs.

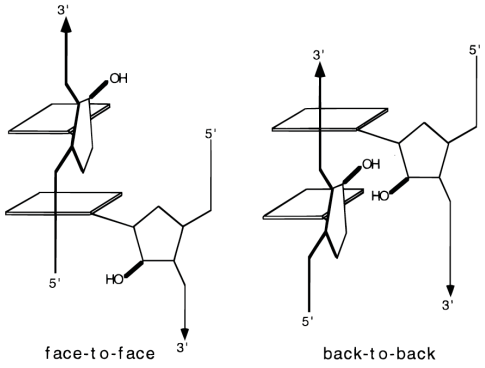


Figure 1.18: Details of the orientation of ribose residues in the narrow groove. Left panel face-to-face and right panel back-to-back orientation. Adapted from reference [263] (American Chemical Society, 1998).

Like RNA, 2'F-RNA also adopts a C3'-*endo* sugar pucker and the fluorine is sterically close to a hydroxyl group. The main difference lies in fluorine's lower hydrogen-bonding capability compared to the 2'-OH group. The introduction of a single 2'-fluorine (**Cf**) in d(TCCCCC) enhances i-motif stability at pH 4.2.[266] Two consecutive modifications, d(TC**Cf**C**f**CC), lead to 0.6 °C reduction compared to the unmodified strand. On the other hand, ribose modifications in the same positions lead to a destabilization of 20 °C. Given the similarity in the size of fluorine and hydroxyl group, the destabilization observed in this case is not due to

steric clashes but likely due to the solvation of the hydroxyl groups compared to a limited solvent accessibility for the fluorinated minor groove. LNA is another RNA mimic locked in the C3'-*endo* conformation with a 2'-O-4'-C methylene bridge. LNA stabilizes DNA and RNA duplexes and triplexes and G4 parallel topologies. LNA was also introduced in the hexamer model sequence dTC_5 at pH 4.0.[267] Some of the partially modified sequences are more stable than the unmodified i-motif, while the fully modified strand does not form an i-motif. The observed stability is attributed to the extended hydrogen-bonding network at the back-to-back steps due to the OCH_2 bridge that extends from C2' to C4'. Therefore, when LNA is carefully introduced, the OCH_2 bridge creates additional hydrogen bonds that neutralize the unfavorable van der Waals contacts in the minor groove.

In conclusion, introducing modifications preorganized in the C3'-*endo* conformation does not necessarily stabilize i-motif structures. On the contrary, most of the C3'-*endo* modifications significantly destabilize or completely abolish i-motif formation due to the presence of electronegative substituents at the 2'-position, leading to steric clashes in the compact narrow groove. Therefore, there is still a need to find modifications that would be well tolerated in the unique i-motif architecture and would result in stabilization at physiological conditions.

1.7.2.2.2 Nucleobase Modifications

Nucleobase modifications have been less investigated compared to sugar and backbone modifications. Wadkins and co-authors indicated that a cytosine modification might have different effects on i-motif stability depending on the environmental conditions.[272] For instance, methylation of cytosine in dilute aqueous solutions leads to an increase in the pK_a and T_m of i-motif structures, while hydroxymethylation leads to a decrease in pK_a and T_m ($\Delta pK_a = -0.2$). In general, halogenated cytosine analogues, such as 5-fluoro, 5-bromo, and 5-iodo, stabilize i-motif formation at acidic pHs,[273] while 5-methylcytosine stabilizes i-motifs at more basic pHs and leads to a slight increase in the pK_a ($\Delta pK_a = +0.11$ for two sequences containing two 5-methylcytosine inserts).[274] Therefore, the pH response of i-motif structures can be tuned by introducing different types of nucleobase modifications.

1.7.2.2.3 Backbone Modifications

The sugar-phosphate backbone of i-motifs is very compact in the minor groove with unusually short distances between adjacent phosphates. Therefore, the i-motif structure is very sensitive to perturbations in the sugar-phosphate backbone. In an attempt to suppress the repulsion between the negatively-charged phosphate backbones, several backbone modifications have been investigated. Molecular dynamics simulations performed on the tetramolecular 5'E and 3'E intercalation topologies of d(CCCC) indicated that the 3'E conformation is more stable due to the presence of two additional sugar interactions along the narrow groove.[275] On the other hand, the 5'E topology of the intramolecular human telomeric i-motif is thermodynamically more stable due to the formation of an additional T-T base pair. Mergny and Lacroix investigated the effect of phosphorothioate, and methylphosphonate, as opposed to the phosphodiester backbone.[276] Their studies suggest that only the phosphodiester and phosphorothioate oligodeoxynucleotides lead to i-motif formation. They hypothesized that the bulkiness of the methyl group prevents i-motif formation in sequences with a methylphosphonate backbone, despite its neutral charge. The incorporation of phosphorothioates in longer sequences (18, 20 and 29 nucleotides) leads to the formation of stable i-motif structures at neutral pH, and they are only a few degrees less stable than the unmodified structures.[276] Additionally, the chirality of the phosphorothioate group influences i-motif stability; for instance the R-stereochemistry leads to greater stabilization compared to the S-stereochemistry (ΔT_m = 11 °C).[277] Another backbone modification that was investigated involved replacing the negatively-charged sugar-phosphate backbone with a neutral polyamide backbone. Peptide nucleic acid (PNA) i-motifs exhibit lower or comparable stability compared to DNA i-motifs.[278] On the contrary, hybrid PNA-DNA i-motifs are more stable than pure PNA or pure DNA i-motifs.[279] This difference can be attributed to the presence of two negatively-charged DNA backbones and two neutral polyamide PNA backbones, rather than a completely neutral backbone or an entirely negatively charged backbone. Another interesting "backbone" modification investigated by Robidoux *et al.* involves branched oligonucleotides, where the 5'-ends of parallel duplexes are joined by a branching riboadenosine linker. Their studies show that branched oligonucleotides can associate into stable i-motif structures.[280]

In Chapter 2, we investigate the effect of 2'F-araC in d(TCCCCC), centromeric, and telomeric sequences. In Chapter 3, we proceed to explore the effect of simultaneously combining

both a sugar modification (2'F-araC) and a nucleobase modification (5-methylcytosine) on the stability and folding topology of i-motif structures. Our rationale for selecting these two modifications is that, among all studied modifications, they stabilize i-motifs at slightly basic pH while increasing the pK_a of the folded structures.

1.7.2.3 The Nature and Length of Connecting Loops

Based on the length of the connecting loops, Brooks *et al.* divided intramolecular i-motif structures into two different classes. i-Motif structures possessing short loops (loop 1 (2-nt) : loop 2 (3/4-nt) : loop 3 (2-nt) *i.e.* 2:3/4:2) were classified as "class I", while "class II" i-motifs possess longer loops (6/8:2/5:6/7).[281] Since longer loops might allow for extra stabilizing interactions, class II i-motifs are considered to be more stable. Very short loops, such as one nucleotide, favor the formation of mono and bimolecular i-motifs, whereas longer loops only lead to the formation of intramolecular i-motif structures.[282] The human telomeric repeat $d(TAACCC)_n$ is stabilized by an A–T base pair between the bases in the loops. This additional base pair stacks on top of the i-motif core, further stabilizing the i-motif structure.[283] Interestingly, d(TCCCGTTTCCA) sequence forms a dimeric i-motif structure *via* the association of two hairpins. This i-motif structure is stabilized up to pH 6.7 due to the presence of T-G-G-T tetrad forming between the two loops.[251] Therefore, longer loops enhance the stabilization of i-motifs only if they can form additional base pairs to further stabilize the i-motif core.

1.7.2.4 Impact of Ionic Strength, Molecular Crowding, and Negative Supercoiling

Ionic strength is another factor that affects the stability of i-motif structures. Unlike G4 structures where the nature of the cation leads to significant differences in stability and folding topology, i-motifs are not affected by the nature of the cation but by the ionic strength of the solution. Mergny and co-workers investigated the effect of salt on the pK_a of cytosine N3 and their results show that low salt concentration reduces it, thereby stabilizing i-motif structures.[282] They showed that increasing NaCl concentration from 0 to 100 mM at a pH close to the pK_a of cytosine destabilizes i-motif structure, while higher NaCl concentrations (300 mM) do not cause further destabilization.[282] The same trend of decreasing i-motif stability with increasing ionic concentration was observed in n-MYC sequences.[284]

Molecular crowding *via* high molecular weight polyethylene glycols (PEG) also contributes to i-motif stability. Molecular crowding mimics the crowded environment in cells, and hence studying its effect on i-motifs is biologically relevant. Crowding conditions preferentially stabilize both i-motif and G4 structures over duplexes and single-stranded sequences.[285] Studies have shown the possible formation of i-motif structures at neutral pH in molecularly-crowded systems which is attributed to the increased pK_a of cytosine N3.[257]

Another factor that is associated with i-motif stability is negative superhelicity. As mentioned in section 1.4, the DNA double helix unwinds into its component single-stranded sequences during transcription. This unwinding leads to negative superhelical stress in the single strands, facilitating the formation of non-canonical secondary structures in the unwound regions.[201] In order to mimic the negative supercoiling induced upstream of a transcription site, Sun and Hurley placed the i-motif and G4-forming sequences of the c-MYC oncogene promoter in a supercoiled plasmid. This study confirmed that negative superhelical stress promotes the formation of i-motifs structures under physiological conditions further highlighting the potential role of i-motifs in regulation of gene expression.[281]

1.7.3 Interaction with Ligands and Proteins

1.7.3.1 Ligands

Compared to the well-documented examples of G4 ligands, the discovery of specific i-motif binding ligands lags far behind. Several ligands have been investigated such as: TMPyP4,[286] bis-acridine (BisA),[287] and phenanthroline compounds.[284] Some of these have a stabilizing effect, but they are not selective to i-motifs and also bind to G4 structures and DNA duplexes. Some metal complexes have been studied as potential i-motif ligands, such as terbium and ruthenium;[288,289] however, they lead to marginal destabilization of i-motif structures and are not i-motif specific. Carboxyl-modified single-walled carbon nanotubes (SWNTs) are considered the first selective i-motif DNA-binding ligands that serve as probes to determine the biological role of i-motif structures *in vivo* and *in vitro* (**Figure 1.19**).

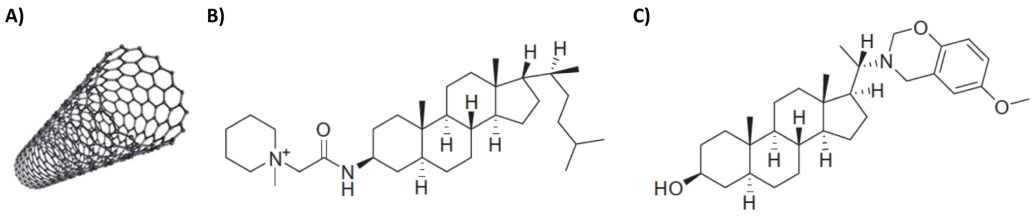


Figure 1.19: Structure of i-motif binding ligands. A) Single-walled carbon nanotubes (SWNT). B) Structure of IMC-48 specific for BCL-2 i-motif. C) Structure of IMC-76 specific for BCL-2 hairpin.

SWNTs were found to inhibit duplex formation in human telomeric regions and induce the formation of i-motif structures at pH 8.0. At pH 5.5, SWNTs lead to a ΔT_m of 22 °C at 10 μg/ml.[290] It is hypothesized that the nanotubes bind to the 5'-end of the major groove in an i-motif, leading to favorable electrostatic interactions between the $C{\cdot}C^+$ core and the carboxylate groups. Qu and co-workers investigated the biomedical effect of SWNTs on telomerase activity and telomere function (discussed in section 1.7.4.1). This study determined that i-motif structures formed in the presence of SWNTs could inhibit telomerase activity, interfere with telomere functions, and lead to senescence and apoptosis in cancer cells *in vitro* and *in vivo*.[291] These experiments hint, but by no means prove, that i-motif structures may play important roles in telomere biology.

Following the carbon nanotubes, only one other study has succeeded in identifying a selective i-motif ligand. After screening a library of 1990 compounds, Hurley and co-workers identified two steroids, compound IMC-48 as an i-motif stabilizing ligand and IMC-76 as a hairpin stabilizing ligand (**Figure 1.19** and **1.20**). This is relevant to the BCL-2 promoter since it exists in dynamic equilibrium between a hairpin and an i-motif structure containing large loops (8:5:7).[292]

IMC-48 binds within the central loop of the i-motif by possible stacking interactions with thymines. On the other hand, the binding site of the lipophilic molecule, IMC-76, was found between the WC hydrogen-bonded regions in the hairpin structure. Furthermore, both of these compounds are selective for BCL-2 over other promoter oncogenes. IMC-48 leads to the activation of gene expression while IMC-76 markedly suppresses the levels of BCL-2 mRNA. It is therefore hypothesized that IMC-48 and IMC-76 are antagonistic and stabilize either the BCL-2 i-motif structure or the hairpin structure, respectively.[292]

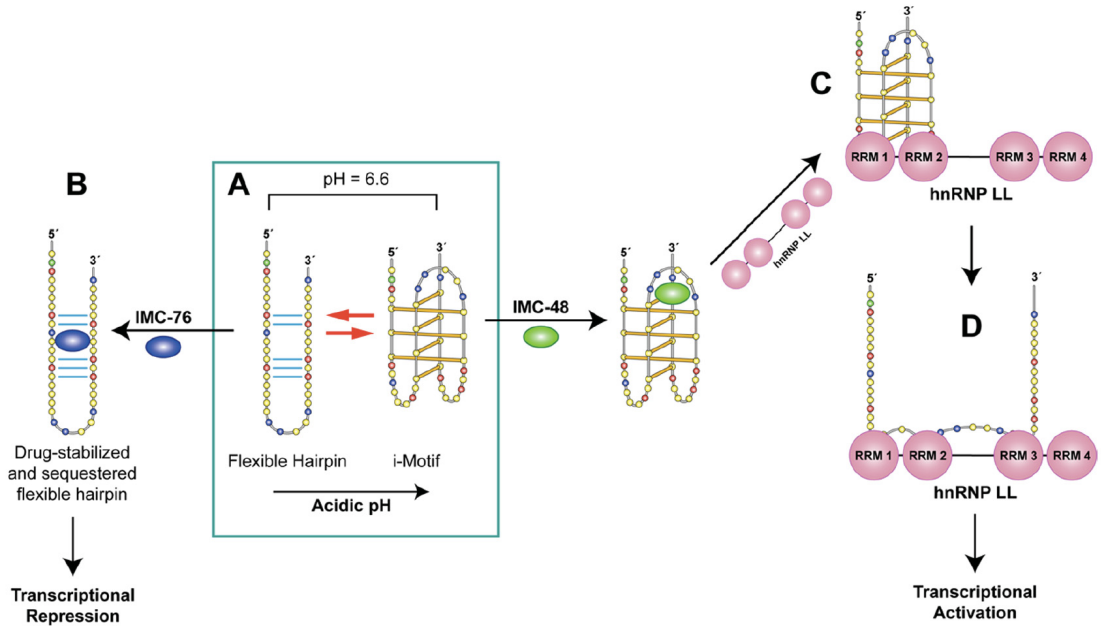


Figure 1.20: A) The conformational equilibrium between the hairpin and the i-motif structure of the C-rich strand in the BCL-2 promoter. B) IMC-76 binds to the hairpin structure leading to transcriptional repression. C) IMC-48 binds to the central loop of the BCL-2 i-motif providing the binding sites for RRMs 1 and 2 of the hnRNP LL. D) hnRNP LL binds to the unfolded C-rich strand causing transcriptional activation of BCL-2. Adapted with permission from reference [293] (American Chemical Society, 2014).

In conclusion, only two i-motif specific ligands have been reported thus far (SWNTs and IMC-48). Most of the other reported ligands are not i-motif specific, and some exhibit higher affinity for G4s and/or DNA duplexes. As more i-motif structures are solved and our understanding of possible binding sites in the compact i-motif structure increases, we will likely see a dramatic increase in the number of i-motif ligands identified. The stabilization of i-motif structures at neutral pH (a goal of this thesis work) will also further our ability to identify ligands that stabilize i-motifs *in vivo*.

1.7.3.2 Proteins

The interaction of C-rich DNA strands with proteins has been investigated; however, in most cases, it is not clear whether the protein binds to a given C-rich strand or to the C-rich strand resulting from i-motif unfolding.[294] Poly-C-binding proteins (PCBP) interact with the C-rich oligonucleotides and play a fundamental role in regulating gene expression. The PCBP family consists of the hnRNP K (heterogeneous nuclear ribonucleoprotein K), αCP1-4, and αCP-KL.[251] In an early study, Lacroix *et al*. investigated the interaction of the hnRNP K and ASF/SF2

proteins with the C-rich telomeric sequence over a wide pH range (6–9.2). Their results show that the protein unfolds i-motif structures at acidic pH and binds to the unfolded C-rich sequence.[294] Additionally, a highly cytosine-specific protein has been discovered in human HeLa cells.[295] This protein is very specific for the human telomeric sequence, $d(CCCTAA)_n$, containing at least four cytosine tracts.[296] However, it is difficult to conclude whether the protein binds to telomeric i-motif structures since the studies were performed at pH 7.4.

The BCL-2 activating transcription factor heterogeneous nuclear ribonucleoprotein LL (hnRNP LL) is one of the very few i-motif binding proteins that have been studied in depth. Through a pull-down assay, Hurley's group identified hnRNP LL from 35 proteins as capable of binding to the i-motif structure of the BCL-2 promoter oncogene and playing a role in transcription.[293] The hnRNP LL protein is a paralog of hnRNP L, which is a premRNA splicing factor capable of binding and stabilizing BCL-2 mRNA.[293] hnRNP LL was found to be specific to i-motif structures as opposed to BCL-2 double-stranded duplex or modified single strand incapable of folding into an i-motif. It is suggested that hnRNP LL binds to the two lateral loops of the BCL-2 i-motif (**Figure 1.20**). CD and bromine footprinting experiments show that the binding of the protein unfolds the i-motif structures to single-stranded sequences. Therefore, the i-motif structures provide the protein binding sites in the most kinetically favorable conformation for binding. Then the protein unfolds the i-motif structures to the more thermodynamically favored single-stranded sequences. The protein remains bound to the single strands and leads to the activation of BCL-2 gene transcription. Therefore, the discovery of the hnRNP LL protein brings i-motif structures into focus as protein recognition sites capable of activating gene expression.

A similar transcription factor, hnRNP K, is known to bind to the C-rich region of the c-MYC promoter.[297] The hnRNP K protein recognizes the TCCC sequences found in the lateral loops of the c-MYC i-motif.[298] Since hnRNP K belongs to the same family of hnRNP LL, it has been hypothesized that hnRNP K activates the transcription of c-MYC gene in a similar way to the activation of BCL-2.

1.7.4 Biological Relevance of i-Motifs *in vivo* – Is There One?

To date, several i-motif forming sequences have been identified and characterized. The recent genomic analysis performed by Burrows and Waller groups confirmed the possible formation of these secondary structures in telomeric, centromeric, and oncogene promoter regions. Some key studies highlighting the impact of i-motif structures in several biological processes are described below.

1.7.4.1 Inhibition of Telomerase Activity

As discussed in section 1.6.5, stabilization of certain human telomeric G-quadruplex topologies with ligands is known to inhibit telomerase activity; however, the role of the complementary C-rich strand has not been investigated in depth. During the S-phase of cell division, the C-rich stand in human telomeres exists transiently as a single-stranded 5'-overhang and therefore has the tendency to fold into i-motif structures.[299] The study conducted by Qu and co-workers in 2012 was the first to investigate the effect of SWNT-stabilized human telomeric i-motifs on telomerase activity (**Figure 1.21**).[291]

SWNTs were found to induce duplex dissociation and to stabilize human telomeric i-motifs under physiological conditions and under molecular crowding conditions. This, in turn, induces the formation of G4 structures on the complementary G-rich strand.[290] In the presence of the i-motif, telomerase activity is inhibited, thereby suggesting that the freed up G4 can no longer be elongated by telomerase. Further investigations into the effect of carboxylated SWNTs on cell growth and telomere structure and function have suggested that the inhibition of cellular growth due to SWNTs was telomere structure-dependent rather than telomerase activity-dependent. This is due to the persistence of i-motif and G4 structures, which lead to telomere uncapping and release of telomere-binding proteins, resulting in telomere dysfunction. Telomere dysfunction induces DNA damage response and activates DNA repair pathways, which in turn trigger cell cycle arrest, senescence, and apoptosis. This report is the first to describe how i-motif structures could inhibit telomerase activity and lead to telomere dysfunction in cancer cells. In Chapter 4, we will investigate the effect of i-motif structures in the telomeric region in the presence of G4 structures that are substrates of telomerase.

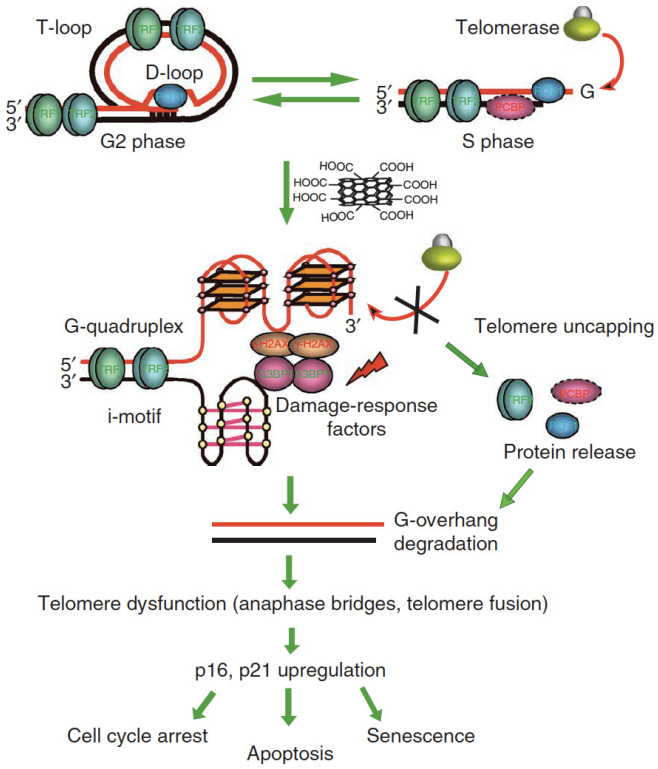


Figure 1.21: The effect of carboxylated SWNTs on telomerase inhibition and telomere uncapping. Adapted from reference [291] (Nature Communication, open access).

1.7.4.2 Transcriptional Regulation of Gene Expression

BCL-2 oncogene promotes cell survival and proliferation through an anti-apoptotic mechanism. It is overexpressed in some cancer cells, while its under-expression leads to neurodegenerative diseases.[300] As discussed earlier, the P1 promoter of the BCL-2 oncogene is in dynamic equilibrium between a hairpin structure and an i-motif structure.[301] Two back-to-back studies by Hurley's group in 2014 showcase small molecules capable of modulating BCL-2 transcription *in vitro* and *in vivo* by specifically targeting either the i-motif or the hairpin form in dynamic equilibrium and *via* the binding of the transcriptional factor (hnRNP LL).[292,293] Stabilization of the i-motif structure *via* IMC-48 leads to significant upregulation of the BCL-2. On the contrary, stabilization of the flexible hairpin species *via* IMC-76 leads to transcriptional repression in breast cancer and lymphoma cell lines *in vitro* and *in vivo* models (**Figure 1.20**).[292] Therefore, these studies demonstrate the antagonism between IMC-48 and IMC-76 in solution and in cell studies.

In a consecutive study, the Hurley group determined the effect of the hnRNP LL protein on BCL-2 transcription. Two of the four RNA recognition motifs (RRMs) in hnRNP LL are required for stable binding to a single-stranded RNA or DNA. The lateral loops of BCL-2 possess sequences capable of binding to the RRMs in hnRNP LL. These studies demonstrate that hnRNP LL activates transcription by recognizing and unfolding the BCL-2 i-motif. The presence of IMC-48 shifts the equilibrium in favor of the i-motif and therefore increases the i-motif population available for binding to hnRNP LL. The research group hypothesizes that IMC-48 binds to the central loop of the BCL-2 i-motif, followed by recognition and binding of hnRNP LL to the two lateral loops. Then, hnRNP LL facilitates i-motif unfolding to bind to the more stable single-stranded sequence. In conclusion, these two studies demonstrate the effect of two small molecules (IMC-76 and IMC-48) and a transcriptional factor (hnRNP LL) on the relative population of i-motif and its impact on gene expression.

1.7.4.3 Regulation of DNA Biosynthesis

Very recently, Sugimoto and co-authors investigated the effect of several non-canonical DNA structures on DNA polymerization by the Klenow fragment (KF) of DNA polymerase.[302] i-Motif structures in the template of a telomere, hypoxia-induced transcription factor, and an insulin-linked polymorphic region (ILPR) were found to stall DNA polymerase and thus impede DNA replication or repair. Despite the fact that i-motifs, hairpin structures, and mixed-type G4s possess similar thermodynamic stabilities, i-motifs were found to be better inhibitors of DNA replication. For instance, KF can easily unwind hairpin structures since they possess a "breathing" terminal base pair (**Figure 1.22**). In the case of mixed G4 structures, the terminal tetrad is more stable than the terminal base pair in a hairpin structure. Therefore, the unzipping of the terminal tetrad is the rate-determining step. Once it is unzipped, the stability of the G4 structure decreases dramatically and polymerase processivity is resumed. On the contrary, the unique intercalating topology of base pairs in i-motif structures complicates its unwinding by polymerase (**Figure 1.22**).

As mentioned earlier, the base pairs in an i-motif structure are intercalated between two parallel duplexes and lack the base stacking interactions like in normal duplexes. Therefore, unzipping is significantly repressed since consecutive base pairs belong to different duplexes. Moreover the loop structure might cause steric hindrance for polymerase binding. Both of these

factors may contribute to the observation that i-motif unwinding by KF requires an activation energy that is threefold higher than that of mixed-G4 unwinding. Therefore, i-motifs can modulate DNA replication to a greater extent than other secondary structures.

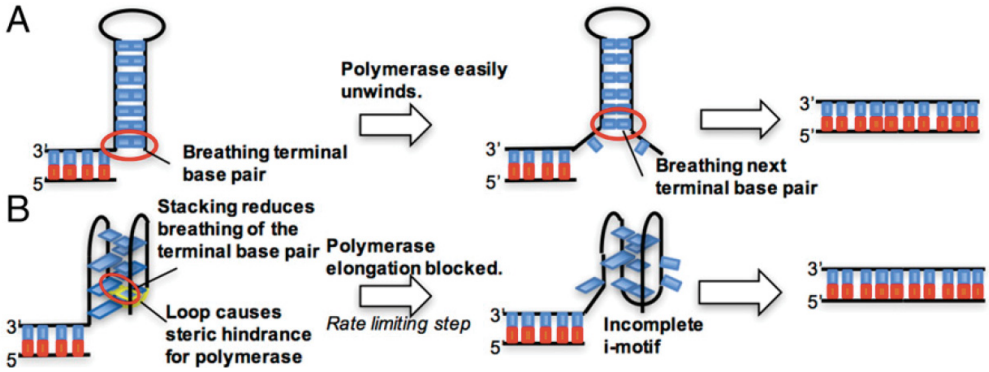


Figure 1.22: The proposed mechanism of unwinding of structured DNAs by the Klenow fragment of DNA polymerase. A) In hairpin structures the terminal base pair is relatively unstable; therefore, polymerase can unzip the hairpin one base pair at a time. B) In i-motif structures, the loops and intercalated base pairs stall the processivity of DNA polymerase. Adapted with permission from reference [302] (National Academy of Science, 2017).

1.7.4.4 Mutual Exclusivity of i-Motifs and G-Quadruplexes and its Effect on Gene Expression

Most of the earlier studies focused on investigating the biological effects of G-quadruplexes and i-motifs separately in single-stranded DNA fragments. However, in cells, these non-canonical secondary structures are flanked between double-stranded DNA duplexes in oncogene promoter regions or present at the ends of chromosomes, which is the case for telomeric sequences. Therefore, the formation of G4 and i-motif structures may compete with duplex DNA, and several factors have been found to tune this equilibrium. Phan and Mergny carried out the first G4/i-motif-duplex interconversion studies. Their results showed that 1:1 mixtures of the G-rich ($AG_3(T_2AG_3)_3$) and C-rich ($((C_3TA_2)_3C_3T$) telomeric sequences at acidic pHs (<5) and in the presence of KCl produced predominantly i-motifs and G4s.[303] However, at pH 7.0 and 100 mM NaCl the duplex was the predominant species even at low temperatures.[304]

Table 1.2: Characteristic features of B-DNA duplexes, G4s, and i-motifs.

Parameters/Characterization	B-DNA duplex	G4	i-motif
Sensitivity to pH	Slight	Slight	Extreme
Sensitivity to cation	No	Yes	No
Molecular crowding	Destabilizes	Stabilizes	Stabilizes
Flexibility	Flexible	Stiff	Very stiff
Selective Ligands	Yes	Yes	Very few
Base pairing	Hetero base pairs AT/GC	Homo base pairs GGGG	Homo base pairs $C{\cdot}C^{+}$
Helical sense	R	R	R
Helical twist	36°	30°	12-16°
Helical rise	3.3-3.4 Å	3.13-3.3 Å	3.1 Å
CD (+ve band, nm)	260-280	Parallel (260) Antiparallel (290)	~ 285
CD (-ve band, nm)	245	Parallel (240) Antiparallel (260)	~ 265
NMR (ppm)	12-14	10-12	15-16

Several groups have investigated the factors influencing interconversion kinetics and determined that the sequence, the experimental conditions (*i.e.* ionic strength, temperature, and pH),[305-307] and the incorporation of chemical modifications (Chapter 4 of the thesis) all play a significant role in favoring the tetraplex structures over duplexes and vice versa. **Table 1.2** lists the factors affecting the stability of G4s, i-motifs, and duplexes along with the distinct CD and NMR signatures resulting from their structural differences.

Recently, several researchers, especially Hurley and Mao, have begun to investigate the co-existence of G4 and i-motif structures. The question as to whether i-motif and G4 structures can co-exist in complementary single strands or are mutually exclusive remains unsolved due to

contradictory reports. One of the earlier reports by Sun and Hurley studied the influence of negative superhelicity on the formation of G4 and i-motif structures in the promoter regions of c-MYC and their impact on gene expression (**Figure 1.23**).[201] Utilizing enzymatic and chemical footprinting, they found that the G4 and i-motif structures co-exist with slight displacement relative to each other. Three out of the four required G·C tracts were shared between the two tetraplexes.

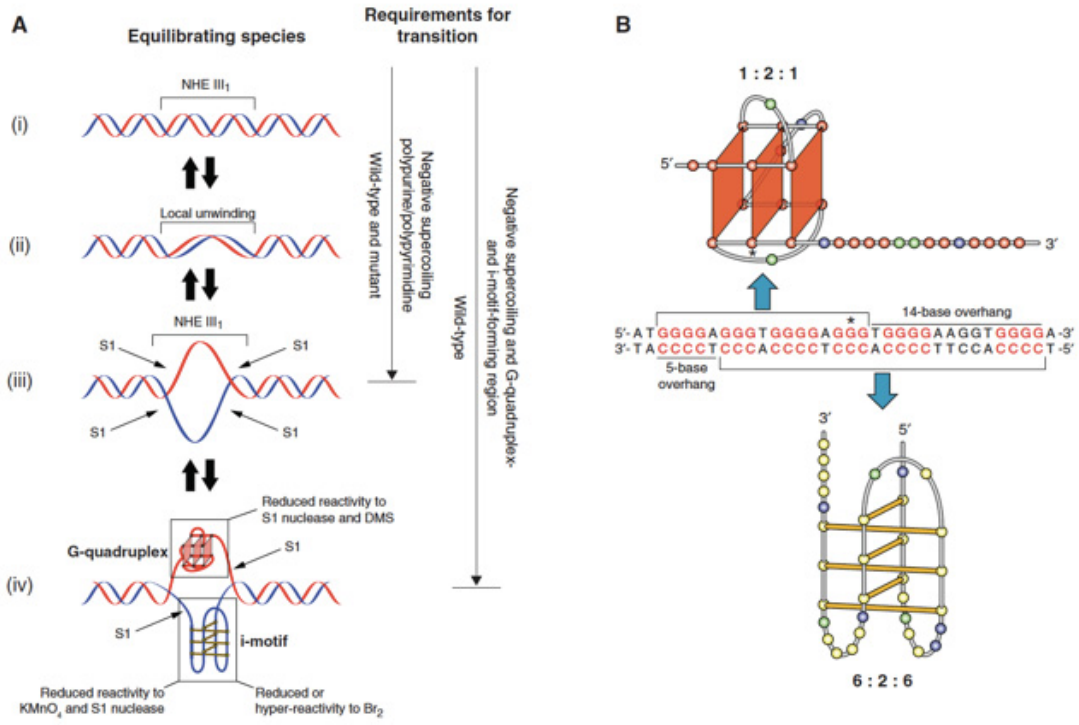


Figure 1.23: A) The proposed equilibrating forms of the NHE III_1 of the c-MYC promoter produced under negative supercoiling. The requirements for the transition from duplexes to single-stranded sequences and then to G4/i-motif species are shown on the right-hand panel. B) The offset positioning of the G4 structure (with a 14 base overhang) and the i-motif structure (with 5-base overhang). Adapted with permission from reference [201] (American Chemical Society, 2009).

On the contrary, a study conducted by Mao's group in 2012 suggests that G4 and i-motif structures are mutually exclusive in a double-stranded DNA in the insulin-linked polymorphic region (ILPR).[308] Using chemical footprinting and laser-tweezers-based single-molecule techniques, they were able to show that the G-rich sequence folds into a G4 structure at pH 7.4 and 100 mM K^+ while the C-rich strand folds into an i-motif structure at pH 5.5 and 100 mM Li^+. Under conditions that favor the formation of both structures (pH 5.5 and 100 mM K^+), either the

G4 or the i-motif structure forms, but not both, probably due to mutual steric hindrance. This observation was further studied in 2016, and results revealed that both tetraplexes form simultaneously when the G4 and i-motif-forming sequences are offset with respect to each other in complementary strands.[309] This study further suggests that mutual exclusivity is governed by steric hindrance in duplex DNA. The mutually exclusive phenomenon also suggests that G4 and i-motif structures may play opposing biological roles at the same location in a double-stranded duplex since G4 formation suppresses transcription, while i-motif structures lead to the activation of gene expression.[281,292]

Following the abovementioned reports, several studies have been reported by Hurley's group to demonstrate how G4s and i-motifs act as on/off molecular switches for the regulation of MYC,[310] platelet-derived growth factor receptor β (PDGFR- β),[311] and KRAS promoters.[312] For instance, in the case of MYC, the SP1 protein induces negative superhelicity freeing up the G- and C-rich strands (**Figure 1.24**). SP1 overexpression leads to a larger melted duplex area, facilitating the binding of hnRNP K to the i-motif structure, thereby activating transcription. On the other hand, nucleolin binds the MYC promoter G4 structure and suppresses transcription. Lastly, in the presence of both nucleolin and hnRNP K, it is apparent that transcription decreases overall. This effect can be attributed to the fact that at low SP1 levels, the G4-nucleolin complex is stronger than the i-motif-hnRNP K complex and therefore the equilibrium is shifted towards the G4-nucleolin complex, which results in repressed transcription.[310]

Despite the lack of direct *in vivo* evidence for the existence of i-motif structures, the examples presented above suggest the likelihood of i-motif formation in biological systems. However, many areas in the i-motif field require further investigations such as the development of specific ligands for i-motifs, the identification of i-motif binding proteins, and the development of i-motif-specific antibodies. The stabilization of i-motif structures under physiological conditions will likely facilitate these endeavors.

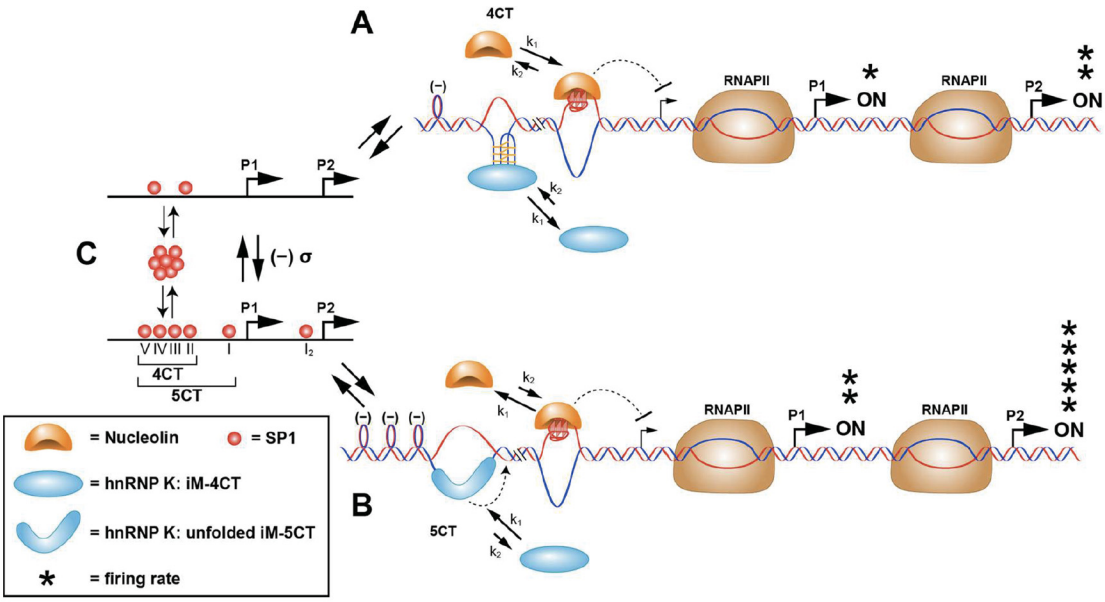


Figure 1.24: The proposed molecular mechanosensor mechanism for the differential control of MYC expression through the NHE III_1. A) Low SP1 levels result in low negative supercoiling, therefore only the 4CT element is accessed, and nucleolin binding to G4 predominates over hnRNP K binding to the i-motif. This results in basal levels of transcription from the P1 and P2 promoters, because hnRNP K can only access the central and lateral loops in the i-motif. B) Higher SP1 levels results in enhanced negative supercoiling, the 5CT element is now available for hnRNP K binding, leading to a more thermodynamically stable complex binding through the addition of the CT element. The binding affinity of hnRNP K to the unfolded i-motif and the additional CT exceeds that of nucleolin to G4, and MYC expression from P1 and P2 is significantly enhanced**.** Adapted with permission from reference [310] (American Chemical Society, 2016).

1.8 Thesis Objectives

The aim of this thesis is to determine the effect of i-motif structures in telomere biology, particularly on the elongation of G4 structures by telomerase. The telomeric G-rich sequence has been studied in depth and several research groups are exploring the role of G4 structures in telomere biology. Certain G4 topologies inhibit telomerase activity, while the parallel intermolecular G4 topology is a substrate of telomerase. On the contrary, the role of the complementary C-rich strand in telomeres has not been well explored yet.

In this thesis, we have attempted to stabilize telomeric i-motif structures at physiological conditions to carry out telomerase assays in the presence of G4 structures. This is key, as we wish to conduct these assays under conditions that will not avert the activity of the enzyme. Therefore, we have introduced chemically-modified nucleoside/nucleotide analogues, especially 2'F-araC, in i-motif structures to help compensate for the factors that typically stabilize i-motif

structures under neutral/physiological pH *in vivo* (*i.e.* negative superhelicity, cellular proteins, and molecular crowding).

In **Chapter 2**, 2'F-araC modifications are introduced in the dTCCCCC hexamer model sequence, in centromeric sequences, and in telomeric sequences capable of forming tetramolecular, bimolecular, and unimolecular i-motif structures, respectively. We investigate the effect of replacing 2-deoxyribose sugar with 2-deoxy-2-fluoro-arabinose *via* thermal melting experiments, circular dichroism, and differential scanning calorimetry. We also report the structural determination of the tetramolecular i-motif structure at pH 5.0 and, for the first time at pH 7.0, *via* 2D NMR analysis.

Building on the results obtained in Chapter 2, **Chapter 3** elaborates on the synthesis of a novel nucleoside analogue (2'-deoxy-2'-fluoro-5-methyl-arabinocytidine) that combines the most stabilizing sugar (2'F-araC) and nucleobase (5-Me-dC) modifications and investigates its impact on the stability of telomeric i-motif structures. Furthermore, we report the first study that combines several cytidine derivatives in the same structure to assess the effects of base and sugar modifications in a "combinatorial" approach. We also provide evidence (*via* NMR, UV, CD) that the stabilization of i-motif structures is highly dependent on the nature of the hemiprotonated base pairs. Lastly, we study the immunostimulatory effect of the novel nucleoside analogue in class C dCpG oligonucleotides.

In **Chapter 4**, first we aim to stabilize parallel intramolecular telomeric G4 structures by incorporating 2'F-araG modifications and determining whether these modified G4 structures are substrates of telomerase *via* direct telomerase activity assays. Next, by incorporating 2'F-araC and 2'F-araG in the C-rich and G-rich telomeric sequences, we were able to monitor and tune the G4/i-motif-duplex interconversion rate at neutral pH. Furthermore, the kinetics of G4/i-motif unfolding and of duplex formation were monitored by 1D NMR experiments. Finally, we performed telomerase activity assays using the constructs that possess stable G4 and i-motif structures simultaneously, in an attempt to determine the effect of i-motif structures on telomerase activity.

1.9 References

1. Blackburn, G. M., Gait M.J., Loakes, D., eds. *Nucleic Acids in Chemistry and Biology*. RSC Publishing, Cambridge, UK, **2006**.
2. Griffith, F., The Significance of Pneumococcal Types. *J. Hyg.* **1928,** *27* (2), 113-59.
3. Avery, O. T.; Macleod, C. M.; McCarty, M., Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii. *J. Exp. Med.* **1944,** *79* (2), 137-58.
4. Magasanik, B.; Vischer, E.; Doniger, R.; Elson, D.; Chargaff, E., The separation and estimation of ribonucleotides in minute quantities. *J. Biol. Chem.* **1950,** *186* (1), 37-50.
5. Franklin, R. E.; Gosling, R. G., Molecular configuration in sodium thymonucleate. *Nature* **1953,** *171* (4356), 740-1.
6. Franklin, R. E.; Gosling, R. G., Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate. *Nature* **1953,** *172* (4369), 156-7.
7. Watson, J. D.; Crick, F. H. C., Molecular Structure of Nucleic Acids - a Structure for Deoxyribose Nucleic Acid. *Nature* **1953,** *171* (4356), 737-738.
8. Watson, J. D.; Crick, F. H. C., Genetical Implications of the Structure of Deoxyribonucleic Acid. *Nature* **1953,** *171* (4361), 964-967.
9. 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), 8205-12.
10. Guschlbauer, W.; Jankowski, K., Nucleoside conformation is determined by the electronegativity of the sugar substituent. *Nucleic Acids Res.* **1980,** *8* (6), 1421-33.
11. Altona, C.; Sundaralingam, M., Conformational analysis of the sugar ring in nucleosides and nucleotides. Improved method for the interpretation of proton magnetic resonance coupling constants. *J. Am. Chem. Soc.* **1973,** *95* (7), 2333-44.
12. Damha, M. J.; Ogilvie, K. K., Conformational Properties of Branched Rna Fragments in Aqueous-Solution. *Biochemistry* **1988,** *27* (17), 6403-6416.
13. Freier, S. M.; Sugimoto, N.; Sinclair, A.; Alkema, D.; Neilson, T.; Kierzek, R.; Caruthers, M. H.; Turner, D. H., Stability of XGCGCp, GCGCYp, and XGCGCYp helixes: an empirical estimate of the energetics of hydrogen bonds in nucleic acids. *Biochemistry* **1986,** *25* (11), 3214-9.
14. Hunter, C. A., Sequence-dependent DNA structure. The role of base stacking interactions. *J. Mol. Biol.* **1993,** *230* (3), 1025-54.
15. Guckian, K. M.; Schweitzer, B. A.; Ren, R. X.; Sheils, C. J.; Paris, P. L.; Tahmassebi, D. C.; Kool, E. T., Experimental Measurement of Aromatic Stacking Affinities in the Context of Duplex DNA. *J. Am. Chem. Soc.* **1996,** *118* (34), 8182-8183.
16. Guckian, K. M.; Schweitzer, B. A.; Ren, R. X.; Sheils, C. J.; Tahmassebi, D. C.; Kool, E. T., Factors Contributing to Aromatic Stacking in Water: Evaluation in the Context of DNA. *J. Am. Chem. Soc.* **2000,** *122* (10), 2213-2222.

17. Sponer, J.; Leszczynski, J.; Hobza, P., Electronic properties, hydrogen bonding, stacking, and cation binding of DNA and RNA bases. *Biopolymers* **2001,** *61* (1), 3-31.
18. Kool, E. T., Hydrogen bonding, base stacking, and steric effects in dna replication. *Annu. Rev. Biophys. Biomol. Struct.* **2001,** *30*, 1-22.
19. Manning, G. S., The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q Rev. Biophys.* **1978,** *11* (2), 179-246.
20. Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E., Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl. Acad. Sci. USA* **1981,** *78* (4), 2179-83.
21. Ghosh, A.; Bansal, M., A glossary of DNA structures from A to Z. *Acta Crystallogr D Biol. Crystallogr.* **2003,** *59* (Pt 4), 620-6.
22. Wang, A. H.; Fujii, S.; van Boom, J. H.; Rich, A., Molecular structure of the octamer d(G-G-C-C-G-G-C-C): modified A-DNA. *Proc. Natl. Acad. Sci. USA* **1982,** *79* (13), 3968-72.
23. Arnott, S.; Chandrasekaran, R.; Millane, R. P.; Park, H. S., RNA-RNA, DNA-DNA, and DNA-RNA Polymorphism. *Biophys. J.* **1986,** *49* (1), 3-5.
24. Saenger, W.; Hunter, W. N.; Kennard, O., DNA Conformation Is Determined by Economics in the Hydration of Phosphate Groups. *Nature* **1986,** *324* (6095), 385-388.
25. Yanagi, K.; Prive, G. G.; Dickerson, R. E., Analysis of Local Helix Geometry in 3 B-DNA Decamers and 8 Dodecamers. *J. Mol. Biol.* **1991,** *217* (1), 201-214.
26. Leslie, A. G. W.; Arnott, S.; Chandrasekaran, R.; Ratliff, R. L., Polymorphism of DNA Double Helices. *J. Mol. Biol.* **1980,** *143* (1), 49-72.
27. Egli, M.; Usman, N.; Zhang, S. G.; Rich, A., Crystal structure of an Okazaki fragment at 2-A resolution. *Proc. Natl. Acad. Sci. USA* **1992,** *89* (2), 534-8.
28. Dervan, P. B., Design of sequence-specific DNA-binding molecules. *Science* **1986,** *232* (4749), 464-71.
29. Zimmer, C.; Wahnert, U., Nonintercalating DNA-binding ligands: specificity of the interaction and their use as tools in biophysical, biochemical and biological investigations of the genetic material. *Prog. Biophys. Mol. Biol.* **1986,** *47* (1), 31-112.
30. Kornberg, A.; Lehman, I. R.; Bessman, M. J.; Simms, E. S., Enzymic Synthesis of Deoxyribonucleic Acid. *Biochim. Biophys. Acta.* **1956,** *21* (1), 197-198.
31. Weiss, S. B.; Nakamoto, T., Net Synthesis of Ribonucleic Acid with a Microbial Enzyme Requiring Deoxyribonucleic Acid and 4 Ribonucleoside Triphosphates. *J. Biol. Chem.* **1961,** *236* (3), Pc18-&.
32. Berget, S. M.; Moore, C.; Sharp, P. A., Spliced Segments at 5' Terminus of Adenovirus 2 Late Messenger-Rna. *Proc. Natl. Acad. Sci. USA* **1977,** *74* (8), 3171-3175.
33. Chow, L. T.; Gelinas, R. E.; Broker, T. R.; Roberts, R. J., Amazing Sequence Arrangement at 5' Ends of Adenovirus-2 Messenger-Rna. *Cell* **1977,** *12* (1), 1-8.
34. Jacob, F.; Monod, J., Genetic Regulatory Mechanisms in Synthesis of Proteins. *J. Mol. Biol.* **1961,** *3* (3), 318.

35. Brenner, S.; Meselson, M.; Jacob, F., Unstabel Intermediate Carrying Information from Genes to Ribosomes for Protein Synthesis. *Nature* **1961,** *190* (477), 576-&.
36. Palade, G. E., A Small Particulate Component of the Cytoplasm. *J. Biophys. Biochem. Cy* **1955,** *1* (1), 59.
37. Cech, T. R.; Zaug, A. J.; Grabowski, P. J., Invitro Splicing of the Ribosomal-Rna Precursor of Tetrahymena - Involvement of a Guanosine Nucleotide in the Excision of the Intervening Sequence. *Cell* **1981,** *27* (3), 487-496.
38. Nissen, P.; Hansen, J.; Ban, N.; Moore, P. B.; Steitz, T. A., The structural basis of ribosome activity in peptide bond synthesis. *Science* **2000,** *289* (5481), 920-930.
39. Hoagland, M. B.; Stephenson, M. L.; Scott, J. F.; Hecht, L. I.; Zamecnik, P. C., Soluble Ribonucleic Acid Intermediate in Protein Synthesis. *J. Biol. Chem.* **1958,** *231* (1), 241-257.
40. Perkins, E. S.; Wood, R. M.; Sears, M. L.; Prusoff, W. H.; Welch, A. D., Anti-viral activities of several iodinated pyrimidine deoxyribonucleosides. *Nature* **1962,** *194*, 985-6.
41. Fyfe, J. A.; Keller, P. M.; Furman, P. A.; Miller, R. L.; Elion, G. B., Thymidine kinase from herpes simplex virus phosphorylates the new antiviral compound, 9-(2-hydroxyethoxymethyl)guanine. *J. Biol. Chem.* **1978,** *253* (24), 8721-7.
42. Robins, R. K., The potential of nucleotide analogs as inhibitors of retroviruses and tumors. *Pharm. Res.* **1984,** *1* (1), 11-8.
43. Jordheim, L. P.; Durantel, D.; Zoulim, F.; Dumontet, C., Advances in the development of nucleoside and nucleotide analogues for cancer and viral diseases. *Nat. Rev. Drug Discov.* **2013,** *12* (6), 447-64.
44. De Clercq, E.; Li, G., Approved Antiviral Drugs over the Past 50 Years. *Clin. Microbiol. Rev* **2016,** *29* (3), 695-747.
45. Shelton, J.; Lu, X.; Hollenbaugh, J. A.; Cho, J. H.; Amblard, F.; Schinazi, R. F., Metabolism, Biochemical Actions, and Chemical Synthesis of Anticancer Nucleosides, Nucleotides, and Base Analogs. *Chem. Rev.* **2016,** *116* (23), 14379-14455.
46. Balzarini, J.; Pauwels, R.; Baba, M.; Herdewijn, P.; de Clercq, E.; Broder, S.; Johns, D. G., The in vitro and in vivo anti-retrovirus activity, and intracellular metabolism of 3'-azido-2',3'-dideoxythymidine and 2',3'-dideoxycytidine are highly dependent on the cell species. *Biochem. Pharmacol.* **1988,** *37* (5), 897-903.
47. Balzarini, J.; Herdewijn, P.; De Clercq, E., Differential patterns of intracellular metabolism of 2',3'-didehydro-2',3'-dideoxythymidine and 3'-azido-2',3'-dideoxythymidine, two potent anti-human immunodeficiency virus compounds. *J. Biol. Chem.* **1989,** *264* (11), 6127-33.
48. Mehellou, Y.; De Clercq, E., Twenty-six years of anti-HIV drug discovery: where do we stand and where do we go? *J. Med. Chem.* **2010,** *53* (2), 521-38.
49. Eriksson, S.; Kierdaszuk, B.; Munch-Petersen, B.; Oberg, B.; Johansson, N. G., Comparison of the substrate specificities of human thymidine kinase 1 and 2 and

deoxycytidine kinase toward antiviral and cytostatic nucleoside analogs. *Biochem. Biophys. Res. Commun.* **1991,** *176* (2), 586-92.

50. Li, Y.; Kong, Y.; Korolev, S.; Waksman, G., Crystal structures of the Klenow fragment of Thermus aquaticus DNA polymerase I complexed with deoxyribonucleoside triphosphates. *Protein Sci.* **1998,** *7* (5), 1116-23.
51. Mu, L.; Sarafianos, S. G.; Nicklaus, M. C.; Russ, P.; Siddiqui, M. A.; Ford, H., Jr.; Mitsuya, H.; Le, R.; Kodama, E.; Meier, C.; Knispel, T.; Anderson, L.; Barchi, J. J., Jr.; Marquez, V. E., Interactions of conformationally biased north and south 2'-fluoro-2', 3'-dideoxynucleoside 5'-triphosphates with the active site of HIV-1 reverse transcriptase. *Biochemistry* **2000,** *39* (37), 11205-15.
52. De Clercq, E., A 40-year journey in search of selective antiviral chemotherapy. *Annu. Rev. Pharmacol. Toxicol.* **2011,** *51*, 1-24.
53. Mitsuya, H.; Weinhold, K. J.; Furman, P. A.; Stclair, M. H.; Lehrman, S. N.; Gallo, R. C.; Bolognesi, D.; Barry, D. W.; Broder, S., 3'-Azido-3'-Deoxythymidine (Bw A509u) - an Antiviral Agent That Inhibits the Infectivity and Cytopathic Effect of Human Lymphotropic-T Virus Type-Iii Lymphadenopathy-Associated Virus Invitro. *Proc. Natl. Acad. Sci. USA* **1985,** *82* (20), 7096-7100.
54. Mitsuya, H.; Broder, S., Inhibition of the Invitro Infectivity and Cytopathic Effect of Human Lymphotrophic-T Virus "Type-Iii/Lymphadenopathy-Associated Virus (Htlv-Iii/Lav) by 2',3'-Dideoxynucleosides. *Proc. Natl. Acad. Sci. USA* **1986,** *83* (6), 1911-1915.
55. Cazenave, C.; Loreau, N.; Thuong, N. T.; Toulme, J. J.; Helene, C., Enzymatic amplification of translation inhibition of rabbit beta-globin mRNA mediated by anti-messenger oligodeoxynucleotides covalently linked to intercalating agents. *Nucleic Acids Res.* **1987,** *15* (12), 4717-36.
56. Mansuri, M. M.; Starrett, J. E., Jr.; Ghazzouli, I.; Hitchcock, M. J.; Sterzycki, R. Z.; Brankovan, V.; Lin, T. S.; August, E. M.; Prusoff, W. H.; Sommadossi, J. P.; et al., 1-(2,3-Dideoxy-beta-D-glycero-pent-2-enofuranosyl)thymine. A highly potent and selective anti-HIV agent. *J. Med. Chem.* **1989,** *32* (2), 461-6.
57. Soudeyns, H.; Yao, X. I.; Gao, Q.; Belleau, B.; Kraus, J. L.; Nguyen-Ba, N.; Spira, B.; Wainberg, M. A., Anti-human immunodeficiency virus type 1 activity and in vitro toxicity of 2'-deoxy-3'-thiacytidine (BCH-189), a novel heterocyclic nucleoside analog. *Antimicrob. Agents Chemother.* **1991,** *35* (7), 1386-1390.
58. Schinazi, R. F.; McMillan, A.; Cannon, D.; Mathis, R.; Lloyd, R. M.; Peck, A.; Sommadossi, J. P.; St Clair, M.; Wilson, J.; Furman, P. A., Selective inhibition of human immunodeficiency viruses by racemates and enantiomers of cis-5-fluoro-1-[2-(hydroxymethyl)-1,3-oxathiolan-5-yl]cytosine. *Antimicrob. Agents Chemother.* **1992,** *36* (11), 2423-2431.
59. Elion, G. B., Mechanism of action and selectivity of acyclovir. *Am. J. Med.* **1982,** *73* (1A), 7-13.

60. Li, F.; Maag, H.; Alfredson, T., Prodrugs of nucleoside analogues for improved oral absorption and tissue targeting. *J. Pharm. Sci.* **2008,** *97* (3), 1109-34.
61. Menendez-Arias, L., Molecular basis of human immunodeficiency virus drug resistance: an update. *Antiviral. Res.* **2010,** *85* (1), 210-31.
62. Steuart, C. D.; Burke, P. J., Cytidine deaminase and the development of resistance to arabinosyl cytosine. *Nat. New Biol.* **1971,** *233* (38), 109-10.
63. Abraham, A.; Varatharajan, S.; Abbas, S.; Zhang, W.; Shaji, R. V.; Ahmed, R.; Abraham, A.; George, B.; Srivastava, A.; Chandy, M.; Mathews, V.; Balasubramanian, P., Cytidine deaminase genetic variants influence RNA expression and cytarabine cytotoxicity in acute myeloid leukemia. *Pharmacogenomics* **2012,** *13* (3), 269-82.
64. Eriksson, S., Is the expression of deoxynucleoside kinases and 5'-nucleotidases in animal tissues related to the biological effects of nucleoside analogs? *Curr. Med. Chem.* **2013,** *20* (34), 4241-8.
65. Hunsucker, S. A.; Mitchell, B. S.; Spychala, J., The 5'-nucleotidases as regulators of nucleotide and drug metabolism. *Pharmacol. Ther.* **2005,** *107* (1), 1-30.
66. Balzarini, J.; Kang, G. J.; Dalal, M.; Herdewijn, P.; De Clercq, E.; Broder, S.; Johns, D. G., The anti-HTLV-III (anti-HIV) and cytotoxic activity of 2',3'-didehydro-2',3'-dideoxyribonucleosides: a comparison with their parental 2',3'-dideoxyribonucleosides. *Mol. Pharmacol.* **1987,** *32* (1), 162-7.
67. Hao, Z.; Cooney, D. A.; Farquhar, D.; Perno, C. F.; Zhang, K.; Masood, R.; Wilson, Y.; Hartman, N. R.; Balzarini, J.; Johns, D. G., Potent DNA chain termination activity and selective inhibition of human immunodeficiency virus reverse transcriptase by 2',3'-dideoxyuridine-5'-triphosphate. *Mol. Pharmacol.* **1990,** *37* (2), 157-63.
68. De Clercq, E.; Holy, A., Acyclic nucleoside phosphonates: a key class of antiviral drugs. *Nat. Rev. Drug. Discov.* **2005,** *4* (11), 928-40.
69. Pradere, U.; Garnier-Amblard, E. C.; Coats, S. J.; Amblard, F.; Schinazi, R. F., Synthesis of nucleoside phosphate and phosphonate prodrugs. *Chem. Rev.* **2014,** *114* (18), 9154-218.
70. Mehellou, Y.; Balzarini, J.; McGuigan, C., Aryloxy phosphoramidate triesters: a technology for delivering monophosphorylated nucleosides and sugars into cells. *ChemMedChem* **2009,** *4* (11), 1779-91.
71. Thornton, P. J.; Kadri, H.; Miccoli, A.; Mehellou, Y., Nucleoside Phosphate and Phosphonate Prodrug Clinical Candidates. *J. Med. Chem.* **2016,** *59* (23), 10400-10410.
72. Mehellou, Y., The ProTides Boom. *ChemMedChem* **2016,** *11* (11), 1114-6.
73. Furman, P. A.; Fyfe, J. A.; St Clair, M. H.; Weinhold, K.; Rideout, J. L.; Freeman, G. A.; Lehrman, S. N.; Bolognesi, D. P.; Broder, S.; Mitsuya, H.; et al., Phosphorylation of 3'-azido-3'-deoxythymidine and selective interaction of the 5'-triphosphate with human immunodeficiency virus reverse transcriptase. *Proc. Natl. Acad. Sci. USA* **1986,** *83* (21), 8333-7.

74. Pertenbreiter, F.; Balzarini, J.; Meier, C., Nucleoside mono- and diphosphate prodrugs of 2',3'-dideoxyuridine and 2',3'-dideoxy-2',3'-didehydrouridine. *ChemMedChem* **2015,** *10* (1), 94-106.
75. Weinschenk, L.; Schols, D.; Balzarini, J.; Meier, C., Nucleoside Diphosphate Prodrugs: Nonsymmetric DiPPro-Nucleotides. *J. Med. Chem.* **2015,** *58* (15), 6114-30.
76. Gollnest, T.; Dinis de Oliveira, T.; Rath, A.; Hauber, I.; Schols, D.; Balzarini, J.; Meier, C., Membrane-permeable Triphosphate Prodrugs of Nucleoside Analogues. *Angewandte Chemie* **2016,** *55* (17), 5255-8.
77. Stephenson, M. L.; Zamecnik, P. C., Inhibition of Rous sarcoma viral RNA translation by a specific oligodeoxyribonucleotide. *Proc. Natl. Acad. Sci. USA* **1978,** *75* (1), 285-8.
78. Zamecnik, P. C.; Stephenson, M. L., Inhibition of Rous sarcoma virus replication and cell transformation by a specific oligodeoxynucleotide. *Proc. Natl. Acad. Sci. USA* **1978,** *75* (1), 280-4.
79. Wiedenheft, B.; Sternberg, S. H.; Doudna, J. A., RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **2012,** *482* (7385), 331-8.
80. Yamamoto, T.; Nakatani, M.; Narukawa, K.; Obika, S., Antisense drug discovery and development. *Future Med. Chem.* **2011,** *3* (3), 339-65.
81. Burnett, J. C.; Rossi, J. J., RNA-based therapeutics: current progress and future prospects. *Chem. Biol.* **2012,** *19* (1), 60-71.
82. Lindow, M.; Kauppinen, S., Discovering the first microRNA-targeted drug. *J. Cell Biol.* **2012,** *199* (3), 407-12.
83. Lennox, K. A.; Behlke, M. A., Chemical modification and design of anti-miRNA oligonucleotides. *Gene Ther.* **2011,** *18* (12), 1111-20.
84. Kole, R.; Krainer, A. R.; Altman, S., RNA therapeutics: beyond RNA interference and antisense oligonucleotides. *Nat. Rev. Drug. Discov.* **2012,** *11* (2), 125-40.
85. Barchet, W.; Wimmenauer, V.; Schlee, M.; Hartmann, G., Accessing the therapeutic potential of immunostimulatory nucleic acids. *Curr. Opin. Immunol.* **2008,** *20* (4), 389-95.
86. Mulhbacher, J.; St-Pierre, P.; Lafontaine, D. A., Therapeutic applications of ribozymes and riboswitches. *Curr. Opin. Pharmacol.* **2010,** *10* (5), 551-6.
87. Chan, C. W.; Khachigian, L. M., DNAzymes and their therapeutic possibilities. *Intern. Med. J.* **2009,** *39* (4), 249-51.
88. Deleavey, G. F.; Damha, M. J., Designing chemically modified oligonucleotides for targeted gene silencing. *Chem. Biol.* **2012,** *19* (8), 937-54.
89. Sharma, V. K.; Watts, J. K., Oligonucleotide therapeutics: chemistry, delivery and clinical progress. *Future Med. Chem.* **2015,** *7* (16), 2221-42.
90. Khvorova, A.; Watts, J. K., The chemical evolution of oligonucleotide therapies of clinical utility. *Nat. Biotechnol.* **2017,** *35* (3), 238-248.
91. Bennett, C. F.; Swayze, E. E., RNA targeting therapeutics: molecular mechanisms of antisense oligonucleotides as a therapeutic platform. *Annu. Rev. Pharmacol. Toxicol.* **2010,** *50*, 259-93.

92. Wu, H.; Lima, W. F.; Zhang, H.; Fan, A.; Sun, H.; Crooke, S. T., Determination of the role of the human RNase H1 in the pharmacology of DNA-like antisense drugs. *J. Biol. Chem.* **2004,** *279* (17), 17181-9.
93. Cerritelli, S. M.; Crouch, R. J., Ribonuclease H: the enzymes in eukaryotes. *FEBS J.* **2009,** *276* (6), 1494-505.
94. Humphreys, D. T.; Westman, B. J.; Martin, D. I.; Preiss, T., MicroRNAs control translation initiation by inhibiting eukaryotic initiation factor 4E/cap and poly(A) tail function. *Proc. Natl. Acad. Sci. USA* **2005,** *102* (47), 16961-6.
95. Matranga, C.; Tomari, Y.; Shin, C.; Bartel, D. P.; Zamore, P. D., Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* **2005,** *123* (4), 607-20.
96. Geary, R. S.; Norris, D.; Yu, R.; Bennett, C. F., Pharmacokinetics, biodistribution and cell uptake of antisense oligonucleotides. *Adv. Drug. Deliv. Rev.* **2015,** *87*, 46-51.
97. Watts, J. K.; Deleavey, G. F.; Damha, M. J., Chemically modified siRNA: tools and applications. *Drug. Discov. Today* **2008,** *13* (19-20), 842-55.
98. Marques, J. T.; Williams, B. R., Activation of the mammalian immune system by siRNAs. *Nat. Biotechnol.* **2005,** *23* (11), 1399-405.
99. Juliano, R. L.; Ming, X.; Nakagawa, O., The chemistry and biology of oligonucleotide conjugates. *Acc. Chem. Res.* **2012,** *45* (7), 1067-76.
100. Lonnberg, H., Solid-phase synthesis of oligonucleotide conjugates useful for delivery and targeting of potential nucleic acid therapeutics. *Bioconjug. Chem.* **2009,** *20* (6), 1065-94.
101. Hair, P.; Cameron, F.; McKeage, K., Mipomersen sodium: first global approval. *Drugs* **2013,** *73* (5), 487-93.
102. Ng, E. W.; Shima, D. T.; Calias, P.; Cunningham, E. T., Jr.; Guyer, D. R.; Adamis, A. P., Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease. *Nat. Rev. Drug. Discov.* **2006,** *5* (2), 123-32.
103. Koshkin, A. A.; Singh, S. K.; Nielsen, P.; Rajwanshi, V. K.; Kumar, R.; Meldgaard, M.; Olsen, C. E.; Wengel, J., LNA (Locked Nucleic Acids): Synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition. *Tetrahedron* **1998,** *54* (14), 3607-3630.
104. Ittig, D.; Liu, S.; Renneberg, D.; Schumperli, D.; Leumann, C. J., Nuclear antisense effects in cyclophilin A pre-mRNA splicing by oligonucleotides: a comparison of tricyclo-DNA with LNA. *Nucleic Acids Res.* **2004,** *32* (1), 346-53.
105. Goyenvalle, A.; Griffith, G.; Babbs, A.; El Andaloussi, S.; Ezzat, K.; Avril, A.; Dugovic, B.; Chaussenot, R.; Ferry, A.; Voit, T.; Amthor, H.; Buhr, C.; Schurch, S.; Wood, M. J.; Davies, K. E.; Vaillend, C.; Leumann, C.; Garcia, L., Functional correction in mouse models of muscular dystrophy using exon-skipping tricyclo-DNA oligomers. *Nat. Med.* **2015,** *21* (3), 270-5.

106. Kool, E. T., Preorganization of DNA: Design Principles for Improving Nucleic Acid Recognition by Synthetic Oligonucleotides. *Chem. Rev.* **1997,** *97* (5), 1473-1488.
107. Mendell, J. R.; Rodino-Klapac, L. R.; Sahenk, Z.; Roush, K.; Bird, L.; Lowes, L. P.; Alfano, L.; Gomez, A. M.; Lewis, S.; Kota, J.; Malik, V.; Shontz, K.; Walker, C. M.; Flanigan, K. M.; Corridore, M.; Kean, J. R.; Allen, H. D.; Shilling, C.; Melia, K. R.; Sazani, P.; Saoud, J. B.; Kaye, E. M.; Eteplirsen Study, G., Eteplirsen for the treatment of Duchenne muscular dystrophy. *Ann. Neurol.* **2013,** *74* (5), 637-47.
108. Garber, K., Big win possible for Ionis/Biogen antisense drug in muscular atrophy. *Nat. Biotechnol.* **2016,** *34* (10), 1002-1003.
109. Watts, J. K.; Sadalapure, K.; Choubdar, N.; Pinto, B. M.; Damha, M. J., Synthesis and conformational analysis of 2 '-fluoro-5-methyl-4 '-thioarabinouridine (4 ' S-FMAU). *J. Org. Chem.* **2006,** *71* (3), 921-925.
110. Trempe, J. F.; Wilds, C. J.; Denisov, A. Y.; Pon, R. T.; Damha, M. J.; Gehring, K., NMR solution structure of an oligonucleotide hairpin with a 2'F-ANA/RNA stem: Implications for RNase H specificity toward DNA/RNA hybrid duplexes. *J. Am. Chem. Soc.* **2001,** *123* (21), 4896-4903.
111. Dowler, T.; Bergeron, D.; Tedeschi, A. L.; Paquet, L.; Ferrari, N.; Damha, M. J., Improvements in siRNA properties mediated by 2'-deoxy-2'-fluoro-beta-D-arabinonucleic acid (FANA). *Nucleic Acids Res.* **2006,** *34* (6), 1669-75.
112. Deleavey, G. F.; Watts, J. K.; Alain, T.; Robert, F.; Kalota, A.; Aishwarya, V.; Pelletier, J.; Gewirtz, A. M.; Sonenberg, N.; Damha, M. J., Synergistic effects between analogs of DNA and RNA improve the potency of siRNA-mediated gene silencing. *Nucleic Acids Res.* **2010,** *38* (13), 4547-4557.
113. Kalota, A.; Karabon, L.; Swider, C. R.; Viazovkina, E.; Elzagheid, M.; Damha, M. J.; Gewirtz, A. M., 2'-deoxy-2'-fluoro-beta-D-arabinonucleic acid (2'F-ANA) modified oligonucleotides (ON) effect highly efficient, and persistent, gene silencing. *Nucleic Acids Res.* **2006,** *34* (2), 451-61.
114. Damha, M. J.; Wilds, C. J.; Noronha, A.; Brukner, I.; Borkow, G.; Arion, D.; Parniak, M. A., Hybrids of RNA and arabinonucleic acids (ANA and 2 ' F-ANA) are substrates of ribonuclease H (vol 120, pg 12976, 1998). *J. Am. Chem. Soc.* **1998,** *120* (51), 13545-13545.
115. Watts, J. K.; Martin-Pintado, N.; Gomez-Pinto, I.; Schwartzentruber, J.; Portella, G.; Orozco, M.; Gonzalez, C.; Damha, M. J., Differential stability of 2 ' F-ANA.RNA and ANA.RNA hybrid duplexes: roles of structure, pseudohydrogen bonding, hydration, ion uptake and flexibility. *Nucleic Acids Res.* **2010,** *38* (7), 2498-2511.
116. Watts, J. K.; Damha, M. J., 2 ' F-arabinonucleic acids (2 ' F-ANA) - History, properties, and new frontiers. *Can. J. Chem.* **2008,** *86* (7), 641-656.
117. Lietard, J.; Abou Assi, H.; Gomez-Pinto, I.; Gonzalez, C.; Somoza, M. M.; Damha, M. J., Mapping the affinity landscape of Thrombin-binding aptamers on 2F-ANA/DNA chimeric G-Quadruplex microarrays. *Nucleic Acids Res.* **2017,** *45* (4), 1619-1632.

118. Istrate, A.; Katolik, A.; Istrate, A.; Leumann, C. J., 2'beta-Fluoro-Tricyclo Nucleic Acids (2'F-tc-ANA): Thermal Duplex Stability, Structural Studies, and RNase H Activation. *Chemistry* **2017,** *23* (43), 10310-10318.
119. Du, Y.; Zhou, X., Targeting non-B-form DNA in living cells. *Chem. Rec.* **2013,** *13* (4), 371-84.
120. Choi, J.; Majima, T., Conformational changes of non-B DNA. *Chem. Soc. Rev.* **2011,** *40* (12), 5893-5909.
121. Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; *et al*. *Nature* **2001,** *409* (6822), 860-921.
122. Wang, A. H.; 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), 680-6.
123. Chakraborty, S.; Sharma, S.; Maiti, P. K.; Krishnan, Y., The poly dA helix: a new structural motif for high performance DNA-based molecular switches. *Nucleic Acids Res.* **2009,** *37* (9), 2810-7.
124. Bonnet, G.; Krichevsky, O.; Libchaber, A., Kinetics of conformational fluctuations in DNA hairpin-loops. *Proc. Natl. Acad. Sci. USA* **1998,** *95* (15), 8602-6.
125. Vasquez, K. M.; Glazer, P. M., Triplex-forming oligonucleotides: principles and applications. *Q Rev. Biophys.* **2002,** *35* (1), 89-107.
126. Brooks, T. A.; Hurley, L. H., The role of supercoiling in transcriptional control of MYC and its importance in molecular therapeutics. *Nat. Rev. Cancer* **2009,** *9* (12), 849-61.
127. Gueron, M.; Leroy, J. L., The i-motif in nucleic acids. *Curr. Opin. Struct. Biol.* **2000,** *10* (3), 326-31.
128. Bacolla, A.; Wells, R. D., Non-B DNA conformations as determinants of mutagenesis and human disease. *Mol. Carcinog.* **2009,** *48* (4), 273-85.
129. Brooks, T. A.; Hurley, L. H., Targeting MYC Expression through G-Quadruplexes. *Genes Cancer* **2010,** *1* (6), 641-649.
130. Wang, G.; Vasquez, K. M., Non-B DNA structure-induced genetic instability. *Mutat. Res.* **2006,** *598* (1-2), 103-19.
131. Cer, R. Z.; Bruce, K. H.; Mudunuri, U. S.; Yi, M.; Volfovsky, N.; Luke, B. T.; Bacolla, A.; Collins, J. R.; Stephens, R. M., Non-B DB: a database of predicted non-B DNA-forming motifs in mammalian genomes. *Nucleic Acids Res.* **2011,** *39* (Database issue), D383-91.
132. Leontis, N. B.; Stombaugh, J.; Westhof, E., The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.* **2002,** *30* (16), 3497-531.
133. Venkatesan, S.; Khaw, A. K.; Hande, M. P., Telomere Biology-Insights into an Intriguing Phenomenon. *Cells* **2017,** *6* (2).
134. Watson, J. D., Origin of concatemeric T7 DNA. *Nat New Biol* **1972,** *239* (94), 197-201.
135. a) Blackburn, E. H.; Epel, E. S.; Lin, J., Human telomere biology: A contributory and interactive factor in aging, disease risks, and protection. *Science* **2015,** *350* (6265), 1193-

8. b) Jayaraman, A.; Kiran, K.G.; Muthusamy, P. *Telomere and Telomerase in Cancer*. InTechOpen, **2016**.
136. Blackburn, E. H.; Gall, J. G., A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. *J. Mol. Biol.* **1978,** *120* (1), 33-53.
137. Blackburn, E. H., Telomere states and cell fates. *Nature* **2000,** *408* (6808), 53-6.
138. Collins, K., Mammalian telomeres and telomerase. *Curr. Opin. Cell Biol.* **2000,** *12* (3), 378-83.
139. Griffith, J. D.; Comeau, L.; Rosenfield, S.; Stansel, R. M.; Bianchi, A.; Moss, H.; de Lange, T., Mammalian telomeres end in a large duplex loop. *Cell* **1999,** *97* (4), 503-14.
140. Oganesian, L.; Bryan, T. M., Physiological relevance of telomeric G-quadruplex formation: a potential drug target. *Bioessays* **2007,** *29* (2), 155-65.
141. De Cian, A.; Lacroix, L.; Douarre, C.; Temime-Smaali, N.; Trentesaux, C.; Riou, J. F.; Mergny, J. L., Targeting telomeres and telomerase. *Biochimie* **2008,** *90* (1), 131-155.
142. Szostak, J. W.; Blackburn, E. H., Cloning yeast telomeres on linear plasmid vectors. *Cell* **1982,** *29* (1), 245-55.
143. Lingner, J.; Hughes, T. R.; Shevchenko, A.; Mann, M.; Lundblad, V.; Cech, T. R., Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* **1997,** *276* (5312), 561-7.
144. Feng, J.; Funk, W. D.; Wang, S. S.; Weinrich, S. L.; Avilion, A. A.; Chiu, C. P.; Adams, R. R.; Chang, E.; Allsopp, R. C.; Yu, J.; et al., The RNA component of human telomerase. *Science* **1995,** *269* (5228), 1236-41.
145. Shay, J. W.; Bacchetti, S., A survey of telomerase activity in human cancer. *Eur. J. Cancer* **1997,** *33* (5), 787-91.
146. Tomlinson, C. G.; Sasaki, N.; Jurczyluk, J.; Bryan, T. M.; Cohen, S. B., Quantitative assays for measuring human telomerase activity and DNA binding properties. *Methods* **2017,** *114*, 85-95.
147. Kim, N. W.; Piatyszek, M. A.; Prowse, K. R.; Harley, C. B.; West, M. D.; Ho, P. L.; Coviello, G. M.; Wright, W. E.; Weinrich, S. L.; Shay, J. W., Specific association of human telomerase activity with immortal cells and cancer. *Science* **1994,** *266* (5193), 2011-5.
148. Wu, Y. Y.; Hruszkewycz, A. M.; Delgado, R. M.; Yang, A.; Vortmeyer, A. O.; Moon, Y. W.; Weil, R. J.; Zhuang, Z.; Remaley, A. T., Limitations on the quantitative determination of telomerase activity by the electrophoretic and ELISA based TRAP assays. *Clin. Chim. Acta.* **2000,** *293* (1-2), 199-212.
149. De Cian, A.; Cristofari, G.; Reichenbach, P.; De Lemos, E.; Monchaud, D.; Teulade-Fichou, M. P.; Shin-Ya, K.; Lacroix, L.; Lingner, J.; Mergny, J. L., Reevaluation of telomerase inhibition by quadruplex ligands and their mechanisms of action. *Proc. Natl. Acad. Sci. USA* **2007,** *104* (44), 17347-52.

150. Yatsunyk, L. A.; Pietrement, O.; Albrecht, D.; Tran, P. L.; Renciuk, D.; Sugiyama, H.; Arbona, J. M.; Aime, J. P.; Mergny, J. L., Guided assembly of tetramolecular G-quadruplexes. *ACS Nano* **2013,** *7* (7), 5701-10.
151. Einarson, O. J.; Sen, D., Self-biotinylation of DNA G-quadruplexes via intrinsic peroxidase activity. *Nucleic Acids Res.* **2017,** *45* (17), 9813-9822.
152. Canale, T. D.; Sen, D., Hemin-utilizing G-quadruplex DNAzymes are strongly active in organic co-solvents. *Biochim. Biophys. Acta.* **2017,** *1861* (5 Pt B), 1455-1462.
153. Neo, J. L.; Kamaladasan, K.; Uttamchandani, M., G-quadruplex based probes for visual detection and sensing. *Curr. Pharm. Des.* **2012,** *18* (14), 2048-57.
154. Neidle, S., Quadruplex Nucleic Acids as Novel Therapeutic Targets. *J. Med. Chem.* **2016,** *59* (13), 5987-6011.
155. Gellert, M.; Lipsett, M. N.; Davies, D. R., Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. USA* **1962,** *48*, 2013-8.
156. Largy, E.; Marchand, A.; Amrane, S.; Gabelica, V.; Mergny, J. L., Quadruplex Turncoats: Cation-Dependent Folding and Stability of Quadruplex-DNA Double Switches. *J. Am. Chem. Soc.* **2016,** *138* (8), 2780-92.
157. Collie, G. W.; Parkinson, G. N., The application of DNA and RNA G-quadruplexes to therapeutic medicines. *Chem. Soc. Rev.* **2011,** *40* (12), 5867-92.
158. Chung, W. J.; Heddi, B.; Schmitt, E.; Lim, K. W.; Mechulam, Y.; Phan, A. T., Structure of a left-handed DNA G-quadruplex. *Proc. Natl. Acad. Sci. USA* **2015,** *112* (9), 2729-33.
159. Sket, P.; Crnugelj, M.; Plavec, J., Identification of mixed di-cation forms of G-quadruplex in solution. *Nucleic Acids Res.* **2005,** *33* (11), 3691-7.
160. Sen, D.; Gilbert, W., A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature* **1990,** *344* (6265), 410-4.
161. Zhang, D.; Huang, T.; Lukeman, P. S.; Paukstelis, P. J., Crystal structure of a DNA/Ba2+ G-quadruplex containing a water-mediated C-tetrad. *Nucleic Acids Res.* **2014,** *42* (21), 13422-9.
162. Reshetnikov, R. V.; Sponer, J.; Rassokhina, O. I.; Kopylov, A. M.; Tsvetkov, P. O.; Makarov, A. A.; Golovin, A. V., Cation binding to 15-TBA quadruplex DNA is a multiple-pathway cation-dependent process. *Nucleic Acids Res.* **2011,** *39* (22), 9789-802.
163. Saintome, C.; Amrane, S.; Mergny, J. L.; Alberti, P., The exception that confirms the rule: a higher-order telomeric G-quadruplex structure more stable in sodium than in potassium. *Nucleic Acids Res.* **2016,** *44* (6), 2926-35.
164. Kim, B. G.; Evans, H. M.; Dubins, D. N.; Chalikian, T. V., Effects of Salt on the Stability of a G-Quadruplex from the Human c-MYC Promoter. *Biochemistry* **2015,** *54* (22), 3420-30.
165. Bugaut, A.; Balasubramanian, S., A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes. *Biochemistry* **2008,** *47* (2), 689-97.

166. Phan, A. T.; Modi, Y. S.; Patel, D. J., Propeller-type parallel-stranded G-quadruplexes in the human c-myc promoter. *J. Am. Chem. Soc.* **2004,** *126* (28), 8710-6.
167. Hazel, P.; Huppert, J.; Balasubramanian, S.; Neidle, S., Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc.* **2004,** *126* (50), 16405-15.
168. Phan, A. T., Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS J.* **2010,** *277* (5), 1107-17.
169. Wang, Y.; Patel, D. J., Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure* **1993,** *1* (4), 263-82.
170. Parkinson, G. N.; Lee, M. P.; Neidle, S., Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **2002,** *417* (6891), 876-80.
171. 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), 15021-7.
172. Ying, L.; Green, J. J.; Li, H.; Klenerman, D.; Balasubramanian, S., Studies on the structure and dynamics of the human telomeric G quadruplex by single-molecule fluorescence resonance energy transfer. *Proc. Natl. Acad. Sci. USA* **2003,** *100* (25), 14629-34.
173. Lee, J. Y.; Okumus, B.; Kim, D. S.; Ha, T., Extreme conformational diversity in human telomeric DNA. *Proc. Natl. Acad. Sci. USA* **2005,** *102* (52), 18938-43.
174. Vorlickova, M.; Chladkova, J.; Kejnovska, I.; Fialova, M.; Kypr, J., Guanine tetraplex topology of human telomere DNA is governed by the number of (TTAGGG) repeats. *Nucleic Acids Res.* **2005,** *33* (18), 5851-60.
175. Harkness, R. W. t.; Mittermaier, A. K., G-register exchange dynamics in guanine quadruplexes. *Nucleic Acids Res.* **2016,** *44* (8), 3481-94.
176. Fleming, A. M.; Zhou, J.; Wallace, S. S.; Burrows, C. J., A Role for the Fifth G-Track in G-Quadruplex Forming Oncogene Promoter Sequences during Oxidative Stress: Do These "Spare Tires" Have an Evolved Function? *ACS Cent. Sci.* **2015,** *1* (5), 226-233.
177. Zhou, J.; Fleming, A. M.; Averill, A. M.; Burrows, C. J.; Wallace, S. S., The NEIL glycosylases remove oxidized guanine lesions from telomeric and promoter quadruplex DNA structures. *Nucleic Acids Res.* **2015,** *43* (8), 4039-54.
178. Lech, C. J.; Cheow Lim, J. K.; Wen Lim, J. M.; Amrane, S.; Heddi, B.; Phan, A. T., Effects of site-specific guanine C8-modifications on an intramolecular DNA G-quadruplex. *Biophys J.* **2011,** *101* (8), 1987-98.
179. Gros, J.; Rosu, F.; Amrane, S.; De Cian, A.; Gabelica, V.; Lacroix, L.; Mergny, J. L., Guanines are a quartet's best friend: impact of base substitutions on the kinetics and stability of tetramolecular quadruplexes. *Nucleic Acids Res.* **2007,** *35* (9), 3064-75.
180. Sacca, B.; Lacroix, L.; Mergny, J. L., The effect of chemical modifications on the thermal stability of different G-quadruplex-forming oligonucleotides. *Nucleic Acids Res.* **2005,** *33* (4), 1182-1192.

181. Peng, C. G.; Damha, M. J., G-quadruplex induced stabilization by 2'-deoxy-2'-fluoro-D-arabinonucleic acids (2'F-ANA). *Nucleic Acids Res.* **2007,** *35* (15), 4977-88.
182. Lech, C. J.; Li, Z.; Heddi, B.; Phan, A. T., 2'-F-ANA-guanosine and 2'-F-guanosine as powerful tools for structural manipulation of G-quadruplexes. *Chem. Commun.* **2012,** *48* (93), 11425-7.
183. Collie, G. W.; Haider, S. M.; Neidle, S.; Parkinson, G. N., A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. *Nucleic Acids Res.* **2010,** *38* (16), 5569-80.
184. Arora, A.; Maiti, S., Differential biophysical behavior of human telomeric RNA and DNA quadruplex. *J. Phys. Chem. B* **2009,** *113* (30), 10515-20.
185. Martin-Pintado, N.; Yahyaee-Anzahaee, M.; Deleavey, G. F.; Portella, G.; Orozco, M.; Damha, M. J.; Gonzalez, C., Dramatic effect of furanose C2' substitution on structure and stability: directing the folding of the human telomeric quadruplex with a single fluorine atom. *J. Am. Chem. Soc.* **2013,** *135* (14), 5344-7.
186. Li, Z.; Lech, C. J.; Phan, A. T., Sugar-modified G-quadruplexes: effects of LNA-, 2'F-RNA- and 2'F-ANA-guanosine chemistries on G-quadruplex structure and stability. *Nucleic Acids Res.* **2014,** *42* (6), 4068-79.
187. Dias, E.; Battiste, J. L.; Williamson, J. R., Chemical Probe for Glycosidic Conformation in Telomeric Dnas. *J. Am. Chem. Soc.* **1994,** *116* (10), 4479-4480.
188. Esposito, V.; Randazzo, A.; Piccialli, G.; Petraccone, L.; Giancola, C.; Mayol, L., Effects of an 8-bromodeoxyguanosine incorporation on the parallel quadruplex structure [d(TGGGT)]4. *Org. Biomol. Chem.* **2004,** *2* (3), 313-8.
189. Dolinnaya, N. G.; Ogloblina, A. M.; Yakubovskaya, M. G., Structure, Properties, and Biological Relevance of the DNA and RNA G-Quadruplexes: Overview 50 Years after Their Discovery. *Biochemistry* **2016,** *81* (13), 1602-1649.
190. Fleming, A. M.; Zhu, J.; Ding, Y.; Burrows, C. J., 8-Oxo-7,8-dihydroguanine in the Context of a Gene Promoter G-Quadruplex Is an On-Off Switch for Transcription. *ACS Chem. Biol.* **2017**, *12*, 2417-2426.
191. Huppert, J. L.; Balasubramanian, S., G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* **2007,** *35* (2), 406-13.
192. Todd, A. K.; Neidle, S., The relationship of potential G-quadruplex sequences in cis-upstream regions of the human genome to SP1-binding elements. *Nucleic Acids Res.* **2008,** *36* (8), 2700-4.
193. Huppert, J. L.; Balasubramanian, S., Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **2005,** *33* (9), 2908-16.
194. Todd, A. K.; Johnston, M.; Neidle, S., Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.* **2005,** *33* (9), 2901-7.
195. Amrane, S.; Adrian, M.; Heddi, B.; Serero, A.; Nicolas, A.; Mergny, J. L.; Phan, A. T., Formation of pearl-necklace monomorphic G-quadruplexes in the human CEB25 minisatellite. *J. Am. Chem. Soc.* **2012,** *134* (13), 5807-16.

196. Agrawal, P.; Lin, C.; Mathad, R. I.; Carver, M.; Yang, D., The major G-quadruplex formed in the human BCL-2 proximal promoter adopts a parallel structure with a 13-nt loop in K+ solution. *J. Am. Chem. Soc.* **2014,** *136* (5), 1750-3.
197. Mukundan, V. T.; Phan, A. T., Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences. *J. Am. Chem. Soc.* **2013,** *135* (13), 5017-28.
198. Beaudoin, J. D.; Perreault, J. P., 5'-UTR G-quadruplex structures acting as translational repressors. *Nucleic Acids Res.* **2010,** *38* (20), 7022-36.
199. Bedrat, A.; Lacroix, L.; Mergny, J. L., Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* **2016,** *44* (4), 1746-59.
200. Chambers, V. S.; Marsico, G.; Boutell, J. M.; Di Antonio, M.; Smith, G. P.; Balasubramanian, S., High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.* **2015,** *33* (8), 877-81.
201. 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.
202. Hatzakis, E.; Okamoto, K.; Yang, D., Thermodynamic stability and folding kinetics of the major G-quadruplex and its loop isomers formed in the nuclease hypersensitive element in the human c-Myc promoter: effect of loops and flanking segments on the stability of parallel-stranded intramolecular G-quadruplexes. *Biochemistry* **2010,** *49* (43), 9152-60.
203. Kuryavyi, V.; Phan, A. T.; Patel, D. J., Solution structures of all parallel-stranded monomeric and dimeric G-quadruplex scaffolds of the human c-kit2 promoter. *Nucleic Acids Res.* **2010,** *38* (19), 6757-73.
204. Islam, B.; Stadlbauer, P.; Krepl, M.; Koca, J.; Neidle, S.; Haider, S.; Sponer, J., Extended molecular dynamics of a c-kit promoter quadruplex. *Nucleic Acids Res.* **2015,** *43* (18), 8673-93.
205. Agrawal, P.; Hatzakis, E.; Guo, K.; Carver, M.; Yang, D., Solution structure of the major G-quadruplex formed in the human VEGF promoter in K+: insights into loop interactions of the parallel G-quadruplexes. *Nucleic Acids Res.* **2013,** *41* (22), 10584-92.
206. Tong, X.; Lan, W.; Zhang, X.; Wu, H.; Liu, M.; Cao, C., Solution structure of all parallel G-quadruplex formed by the oncogene RET promoter sequence. *Nucleic Acids Res.* **2011,** *39* (15), 6753-63.
207. Waller, Z. A.; Howell, L. A.; Macdonald, C. J.; O'Connell, M. A.; Searcey, M., Identification and characterisation of a G-quadruplex forming sequence in the promoter region of nuclear factor (erythroid-derived 2)-like 2 (Nrf2). *Biochem. Biophys. Res. Commun.* **2014,** *447* (1), 128-32.
208. Rodriguez, R.; Miller, K. M.; Forment, J. V.; Bradshaw, C. R.; Nikan, M.; Britton, S.; Oelschlaegel, T.; Xhemalce, B.; Balasubramanian, S.; Jackson, S. P., Small-molecule-

induced DNA damage identifies alternative DNA structures in human genes. *Nat. Chem. Biol.* **2012,** *8* (3), 301-10.

209. Chaires, J. B.; Trent, J. O.; Gray, R. D.; Dean, W. L.; Buscaglia, R.; Thomas, S. D.; Miller, D. M., An improved model for the hTERT promoter quadruplex. *Plos One* **2014,** *9* (12), e115580.
210. Murat, P.; Zhong, J.; Lekieffre, L.; Cowieson, N. P.; Clancy, J. L.; Preiss, T.; Balasubramanian, S.; Khanna, R.; Tellam, J., G-quadruplexes regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation. *Nat. Chem. Biol.* **2014,** *10* (5), 358-64.
211. Fleming, A. M.; Ding, Y.; Alenko, A.; Burrows, C. J., Zika Virus Genomic RNA Possesses Conserved G-Quadruplexes Characteristic of the Flaviviridae Family. *ACS Infect. Dis.* **2016,** *2* (10), 674-681.
212. Fernando, H.; Rodriguez, R.; Balasubramanian, S., Selective recognition of a DNA G-quadruplex by an engineered antibody. *Biochemistry* **2008,** *47* (36), 9365-9371.
213. Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S., Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* **2013,** *5* (3), 182-6.
214. 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.
215. Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H., Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. USA* **2002,** *99* (18), 11593-8.
216. Cogoi, S.; Xodo, L. E., G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.* **2006,** *34* (9), 2536-49.
217. David, A. P.; Margarit, E.; Domizi, P.; Banchio, C.; Armas, P.; Calcaterra, N. B., G-quadruplexes as novel cis-elements controlling transcription during embryonic development. *Nucleic Acids Res.* **2016,** *44* (9), 4163-4173.
218. Seenisamy, J.; Rezler, E. M.; Powell, T. J.; Tye, D.; Gokhale, V.; Joshi, C. S.; Siddiqui-Jain, A.; Hurley, L. H., The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4. *J. Am. Chem. Soc.* **2004,** *126* (28), 8702-9.
219. Palumbo, S. L.; Memmott, R. M.; Uribe, D. J.; Krotova-Khan, Y.; Hurley, L. H.; Ebbinghaus, S. W., A novel G-quadruplex-forming GGA repeat region in the c-myb promoter is a critical regulator of promoter activity. *Nucleic Acids Res.* **2008,** *36* (6), 1755-69.
220. Johnson, J. E.; Cao, K.; Ryvkin, P.; Wang, L. S.; Johnson, F. B., Altered gene expression in the Werner and Bloom syndromes is associated with sequences having G-quadruplex forming potential. *Nucleic Acids Res.* **2010,** *38* (4), 1114-22.
221. Nguyen, G. H.; Tang, W.; Robles, A. I.; Beyer, R. P.; Gray, L. T.; Welsh, J. A.; Schetter, A. J.; Kumamoto, K.; Wang, X. W.; Hickson, I. D.; Maizels, N.; Monnat, R. J., Jr.; Harris,

C. C., Regulation of gene expression by the BLM helicase correlates with the presence of G-quadruplex DNA motifs. *Proc. Natl. Acad. Sci. USA* **2014,** *111* (27), 9905-10.
222. Smith, J. S.; Chen, Q.; Yatsunyk, L. A.; Nicoludis, J. M.; Garcia, M. S.; Kranaster, R.; Balasubramanian, S.; Monchaud, D.; Teulade-Fichou, M. P.; Abramowitz, L.; Schultz, D. C.; Johnson, F. B., Rudimentary G-quadruplex-based telomere capping in Saccharomyces cerevisiae. *Nat. Struct. Mol. Biol.* **2011,** *18* (4), 478-85.
223. Yatsunyk, L. A.; Bryan, T. M.; Johnson, F. B., G-ruption: the third international meeting on G-quadruplex and G-assembly. *Biochimie* **2012,** *94* (12), 2475-83.
224. Neidle, S., Human telomeric G-quadruplex: the current status of telomeric G-quadruplexes as therapeutic targets in human cancer. *FEBS J.* **2010,** *277* (5), 1118-25.
225. Zahler, A. M.; Williamson, J. R.; Cech, T. R.; Prescott, D. M., Inhibition of Telomerase by G-Quartet DNA Structures. *Nature* **1991,** *350* (6320), 718-720.
226. Oganesian, L.; Moon, I. K.; Bryan, T. M.; Jarstfer, M. B., Extension of G-quadruplex DNA by ciliate telomerase. *EMBO J.* **2006,** *25* (5), 1148-59.
227. Moye, A. L.; Porter, K. C.; Cohen, S. B.; Phan, T.; Zyner, K. G.; Sasaki, N.; Lovrecz, G. O.; Beck, J. L.; Bryan, T. M., Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nat. Commun.* **2015,** *6*, 1-12.
228. Hampel, S. M.; Sidibe, A.; Gunaratnam, M.; Riou, J. F.; Neidle, S., Tetrasubstituted naphthalene diimide ligands with selectivity for telomeric G-quadruplexes and cancer cells. *Bioorg. Med. Chem. Lett.* **2010,** *20* (22), 6459-63.
229. Chung, W. J.; Heddi, B.; Tera, M.; Iida, K.; Nagasawa, K.; Phan, A. T., Solution structure of an intramolecular (3 + 1) human telomeric G-quadruplex bound to a telomestatin derivative. *J. Am. Chem. Soc.* **2013,** *135* (36), 13495-501.
230. Burger, A. M.; Dai, F.; Schultes, C. M.; Reszka, A. P.; Moore, M. J.; Double, J. A.; Neidle, S., The G-quadruplex-interactive molecule BRACO-19 inhibits tumor growth, consistent with telomere targeting and interference with telomerase function. *Cancer Res.* **2005,** *65* (4), 1489-96.
231. Phatak, P.; Cookson, J. C.; Dai, F.; Smith, V.; Gartenhaus, R. B.; Stevens, M. F.; Burger, A. M., Telomere uncapping by the G-quadruplex ligand RHPS4 inhibits clonogenic tumour cell growth in vitro and in vivo consistent with a cancer stem cell targeting mechanism. *Br. J. Cancer* **2007,** *96* (8), 1223-33.
232. Tauchi, T.; Shin-ya, K.; Sashida, G.; Sumi, M.; Okabe, S.; Ohyashiki, J. H.; Ohyashiki, K., Telomerase inhibition with a novel G-quadruplex-interactive agent, telomestatin: in vitro and in vivo studies in acute leukemia. *Oncogene* **2006,** *25* (42), 5719-25.
233. Biffi, G.; Tannahill, D.; Balasubramanian, S., An intramolecular G-quadruplex structure is required for binding of telomeric repeat-containing RNA to the telomeric protein TRF2. *J. Am. Chem. Soc.* **2012,** *134* (29), 11974-6.
234. Chakraborty, P.; Grosse, F., Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA repair* **2011,** *10* (6), 654-65.

235. Li, Q. J.; Tong, X. J.; Duan, Y. M.; Zhou, J. Q., Characterization of the intramolecular G-quadruplex promoting activity of Est1. *Febs Lett.* **2013,** *587* (6), 659-65.
236. Paeschke, K.; Bochman, M. L.; Garcia, P. D.; Cejka, P.; Friedman, K. L.; Kowalczykowski, S. C.; Zakian, V. A., Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature* **2013,** *497* (7450), 458-62.
237. Gray, L. T.; Vallur, A. C.; Eddy, J.; Maizels, N., G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD. *Nat. Chem. Biol.* **2014,** *10* (4), 313-8.
238. Gonzalez, V.; Guo, K. X.; Hurley, L.; Sun, D., Identification and Characterization of Nucleolin as a c-myc G-quadruplex-binding Protein. *J. Biol. Chem.* **2009,** *284* (35), 23622-23635.
239. Kang, H. J.; Le, T. V.; Kim, K.; Hur, J.; Kim, K. K.; Park, H. J., Novel interaction of the Z-DNA binding domain of human ADAR1 with the oncogenic c-Myc promoter G-quadruplex. *J. Mol. Biol.* **2014,** *426* (14), 2594-604.
240. Marsh, R. E.; Bierstedt, R.; Eichhorn, E. L., Crystal Structure of Cytosine-5-Acetic Acid. *Acta. Crystallogr.* **1962,** *15* (Apr), 310-&.
241. Gehring, K.; Leroy, J. L.; Gueron, M., A Tetrameric DNA-Structure with Protonated Cytosine.Cytosine Base-Pairs. *Nature* **1993,** *363* (6429), 561-565.
242. Alberti, P.; Bourdoncle, A.; Sacca, B.; Lacroix, L.; Mergny, J. L., DNA nanomachines and nanostructures involving quadruplexes. *Org. Biomol. Chem.* **2006,** *4* (18), 3383-91.
243. Li, T.; Ackermann, D.; Hall, A. M.; Famulok, M., Input-dependent induction of oligonucleotide structural motifs for performing molecular logic. *J. Am. Chem. Soc.* **2012,** *134* (7), 3508-16.
244. Narayanaswamy, N.; Nair, R. R.; Suseela, Y. V.; Saini, D. K.; Govindaraju, T., A molecular beacon-based DNA switch for reversible pH sensing in vesicles and live cells. *Chem. Commun.* **2016,** *52* (56), 8741-4.
245. Surana, S.; Bhat, J. M.; Koushika, S. P.; Krishnan, Y., An autonomous DNA nanomachine maps spatiotemporal pH changes in a multicellular living organism. *Nat. Commun.* **2011,** *2*, 340.
246. Kang, C. H.; Berger, I.; Lockshin, C.; Ratliff, R.; Moyzis, R.; Rich, A., Crystal-Structure of Intercalated 4-Stranded D(C3t) at 1.4 Angstrom Resolution. *Proc. Natl. Acad. Sci. USA* **1994,** *91* (24), 11636-11640.
247. 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), 282-90.
248. Lieblein, A. L.; Kramer, M.; Dreuw, A.; Furtig, B.; Schwalbe, H., The nature of hydrogen bonds in cytidine...H+...cytidine DNA base pairs. *Angew. Chem.* **2012,** *51* (17), 4067-70.
249. Leroy, J. L.; Gehring, K.; Kettani, A.; Gueron, M., Acid multimers of oligodeoxycytidine strands: stoichiometry, base-pair characterization, and proton exchange properties. *Biochemistry* **1993,** *32* (23), 6019-31.

250. Phan, A. T.; Gueron, 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.
251. Benabou, S.; Aviñó, A.; Eritja, R.; González, C.; Gargallo, R., Fundamental aspects of the nucleic acid i-motif structures. *Rsc Adv.* **2014,** *4* (51), 26956.
252. Berger, I.; Egli, M.; Rich, A., Inter-strand C-H···O hydrogen bonds stabilizing four-stranded intercalated molecules: Stereoelectronic effects of 04' in cytosine-rich DNA. *Proc. Natl. Acad. Sci. USA* **1996,** *93* (22), 12116-12121.
253. Berger, I.; Kang, C.; Fredian, A.; Ratliff, R.; Moyzis, R.; Rich, A., Extension of the 4-Stranded Intercalated Cytosine Motif by Adenine-Center-Dot-Adenine Base-Pairing in the Crystal-Structure of D(Cccaat). *Nat. Struct. Biol.* **1995,** *2* (5), 416-429.
254. Phan, A. T.; Leroy, J. L., Intramolecular i-motif structures of telomeric DNA. *J. Biomol. Struct. Dyn.* **2000**, 245-251.
255. Lieblein, A. L.; Buck, J.; Schlepckow, K.; Furtig, B.; Schwalbe, H., Time-resolved NMR spectroscopic studies of DNA i-motif folding reveal kinetic partitioning. *Angew. Chem.* **2012,** *51* (1), 250-3.
256. Zhou, J.; Wei, C.; Jia, G.; Wang, X.; Feng, Z.; Li, C., Formation of i-motif structure at neutral and slightly alkaline pH. *Molecular bioSystems* **2010,** *6* (3), 580-6.
257. Rajendran, A.; Nakano, S.; Sugimoto, N., Molecular crowding of the cosolutes induces an intramolecular i-motif structure of triplet repeat DNA oligomers at neutral pH. *Chem. Commun.* **2010,** *46* (8), 1299-301.
258. Day, H. A.; Huguin, C.; Waller, Z. A. E., Silver cations fold i-motif at neutral pH. *Chem. Commun.* **2013,** *49* (70), 7696-7698.
259. Fojtik, P.; Vorlickova, M., The fragile X chromosome (GCC) repeat folds into a DNA tetraplex at neutral pH. *Nucleic Acids Res.* **2001,** *29* (22), 4684-90.
260. 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.
261. 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.
262. 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), e11647.
263. 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.
264. Lacroix, L.; Mergny, J. L.; Leroy, J. L.; Helene, C., Inability of RNA to form the i-motif: Implications for triplex formation. *Biochemistry* **1996,** *35* (26), 8715-8722.

265. 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), 529-35.
266. Fenna, C. P.; Wilkinson, V. J.; Arnold, J. R.; Cosstick, R.; Fisher, J., The effect of 2'-fluorine substitutions on DNA i-motif conformation and stability. *Chem. Commun.* **2008,** (30), 3567-9.
267. Kumar, N.; Nielsen, J. T.; Maiti, S.; Petersen, M., i-Motif formation with locked nucleic acid (LNA). *Angew. Chem.* **2007,** *46* (48), 9220-2.
268. 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), 752-5.
269. Perez-Rentero, S.; Gargallo, R.; Gonzalez, C.; Eritja, R., Modulation of the stability of i-motif structures using an acyclic threoninol cytidine derivative. *Rsc Adv.* **2015,** *5* (78), 63278-63281.
270. Abou Assi, H.; Harkness, R. W. V.; Martin-Pintado, N.; Wilds, C. J.; Campos-Olivas, R.; Mittermaier, A. K.; Gonzalez, C.; Damha, M. J., Stabilization of i-motif structures by 2'-beta-fluorination of DNA. *Nucleic Acids Res.* **2016,** *44* (11), 4998-5009.
271. Snoussi, K.; Nonin-Lecomte, S.; Leroy, J. L., The RNA i-motif. *J. Mol. Biol.* **2001,** *309* (1), 139-153.
272. Bhaysar-Jog, Y. P.; Van Dornshuld, E.; Brooks, T. A.; Tschumper, G. S.; Wadkins, R. M., Epigenetic Modification, Dehydration, and Molecular Crowding Effects on the Thermodynamics of i-Motif Structure Formation from C-Rich DNA. *Biochemistry* **2014,** *53* (10), 1586-1594.
273. Lannes, L.; Halder, S.; Krishnan, Y.; Schwalbe, H., Tuning the pH Response of i-Motif DNA Oligonucleotides. *Chembiochem* **2015,** *16* (11), 1647-1656.
274. Xu, B. C.; Devi, G.; Shao, F. W., Regulation of telomeric i-motif stability by 5-methylcytosine and 5-hydroxymethylcytosine modification. *Org. Biomol. Chem.* **2015,** *13* (20), 5646-5651.
275. 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.
276. Mergny, J. L.; Lacroix, L., Kinetics and thermodynamics of i-DNA formation: phosphodiester versus modified oligodeoxynucleotides. *Nucleic Acids Res.* **1998,** *26* (21), 4797-4803.
277. Kanaori, K.; Sakamoto, S.; Yoshida, H.; Guga, P.; Stec, W.; Tajima, K.; Makino, K., Effect of phosphorothioate chirality on i-motif structure and stability. *Biochemistry* **2004,** *43* (19), 5672-5679.
278. Volker, J.; Klump, H. H.; Breslauer, K. J., The energetics of i-DNA tetraplex structures formed intermolecularly by d(TC5) and intramolecularly by d[(C5T3)3C5]. *Biopolymers* **2007,** *86* (2), 136-47.

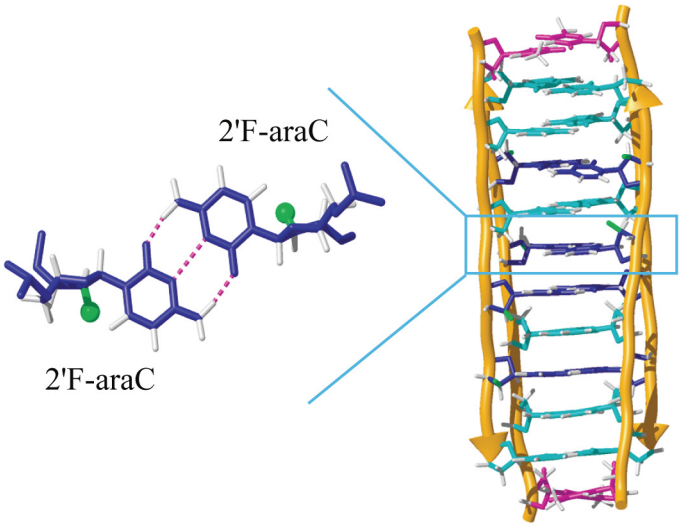
279. Mathur, V.; Verma, A.; Maiti, S.; Chowdhury, S., Thermodynamics of i-tetraplex formation in the nuclease hypersensitive element of human c-myc promoter. *Biochem. Biophys. Res. Commun.* **2004,** *320* (4), 1220-7.
280. Robidoux, S.; Klinck, R.; Gehring, K.; Damha, M. J., Association of branched oligonucleotides into the i-motif. *J. Biomol. Struct. Dyn.* **1997,** *15* (3), 517-27.
281. Brooks, T. A.; Kendrick, S.; Hurley, L., Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS J* **2010,** *277* (17), 3459-69.
282. Mergny, J. L.; Lacroix, L.; Han, X. G.; Leroy, J. L.; Helene, C., Intramolecular Folding of Pyrimidine Oligodeoxynucleotides into an I-DNA Motif. *J. Am. Chem. Soc.* **1995,** *117* (35), 8887-8898.
283. Kang, C. H.; Berger, I.; Lockshin, C.; Ratliff, R.; Moyzis, R.; Rich, A., Stable Loop in the Crystal-Structure of the Intercalated 4-Stranded Cytosine-Rich Metazoan Telomere. *Proc. Natl. Acad. Sci. USA* **1995,** *92* (9), 3874-3878.
284. Day, H. A.; Pavlou, P.; Waller, Z. A., i-Motif DNA: structure, stability and targeting with ligands. *Bioorg. Med. Chem.* **2014,** *22* (16), 4407-18.
285. 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), 12751-67.
286. Martino, L.; Pagano, B.; Fotticchia, I.; Neidle, S.; Giancola, C., Shedding light on the interaction between TMPyP4 and human telomeric quadruplexes. *J. Phys. Chem. B* **2009,** *113* (44), 14779-86.
287. Alberti, P.; Ren, J.; Teulade-Fichou, M. P.; Guittat, L.; Riou, J. F.; Chaires, J.; Helene, C.; Vigneron, J. P.; Lehn, J. M.; Mergny, J. L., Interaction of an acridine dimer with DNA quadruplex structures. *J. Biomol. Struct. Dyn.* **2001,** *19* (3), 505-13.
288. Xu, H.; Zhang, H.; Qu, X., Interactions of the human telomeric DNA with terbium-amino acid complexes. *J. Inorg. Biochem.* **2006,** *100* (10), 1646-52.
289. Shi, S.; Geng, X.; Zhao, J.; Yao, T.; Wang, C.; Yang, D.; Zheng, L.; Ji, L., Interaction of [Ru(bpy)2(dppz)]2+ with human telomeric DNA: preferential binding to G-quadruplexes over i-motif. *Biochimie* **2010,** *92* (4), 370-7.
290. 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. USA* **2006,** *103* (52), 19658-63.
291. 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*, 1074.
292. Kendrick, S.; Kang, H. J.; Alam, M. P.; Madathil, M. M.; Agrawal, P.; Gokhale, V.; Yang, D. Z.; 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.

293. 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. *J. Am. Chem. Soc.* **2014,** *136* (11), 4172-4185.
294. Lacroix, L.; Lienard, H.; Labourier, E.; Djavaheri-Mergny, M.; Lacoste, J.; Leffers, H.; Tazi, J.; Helene, C.; Mergny, J. L., Identification of two human nuclear proteins that recognise the cytosine-rich strand of human telomeres in vitro. *Nucleic Acids Res.* **2000,** *28* (7), 1564-1575.
295. Marsich, E.; Piccini, A.; Xodo, L. E.; Manzini, G., Evidence for a HeLa nuclear protein that binds specifically to the single-stranded d(CCCTAA)(n) telomeric motif. *Nucleic Acids Res.* **1996,** *24* (20), 4029-4033.
296. Marsich, E.; Xodo, L. E.; Manzini, G., Widespread presence in mammals and high binding specificity of a nuclear protein that recognises the single-stranded telomeric motif (CCCTAA)n. *Eur. J. Biochem.* **1998,** *258* (1), 93-9.
297. Gonzalez, V.; Hurley, L. H., The c-MYC NHE III(1): function and regulation. *Annu. Rev. Pharmacol. Toxicol.* **2010,** *50*, 111-29.
298. Backe, P. H.; Messias, A. C.; Ravelli, R. B.; Sattler, M.; Cusack, S., X-ray crystallographic and NMR studies of the third KH domain of hnRNP K in complex with single-stranded nucleic acids. *Structure* **2005,** *13* (7), 1055-67.
299. Cimino-Reale, G.; Pascale, E.; Alvino, E.; Starace, G.; D'Ambrosio, E., Long telomeric C-rich 5'-tails in human replicating cells. *J. Biol. Chem.* **2003,** *278* (4), 2136-40.
300. Bar-Am, O.; Weinreb, O.; Amit, T.; Youdim, M. B., Regulation of Bcl-2 family proteins, neurotrophic factors, and APP processing in the neurorescue activity of propargylamine. *Faseb J* **2005,** *19* (13), 1899-901.
301. 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.
302. Takahashi, S.; Brazier, J. A.; Sugimoto, N., Topological impact of noncanonical DNA structures on Klenow fragment of DNA polymerase. *Proc. Natl. Acad. Sci. USA* **2017,** *114* (36), 9605-9610.
303. Phan, A. T.; Mergny, J. L., Human telomeric DNA: G-quadruplex, i-motif and watson-crick double helix. *Nucleic Acids Res.* **2002,** *30* (21), 4618-4625.
304. Li, W.; Wu, P.; Ohmichi, T.; Sugimoto, N., Characterization and thermodynamic properties of quadruplex/duplex competition. *Febs Lett.* **2002,** *526* (1-3), 77-81.
305. Bucek, P.; Jaumot, J.; Avino, A.; Eritja, R.; Gargallo, R., pH-Modulated Watson-Crick Duplex-Quadruplex Equilibria of Guanine-Rich and Cytosine-Rich DNA Sequences 140 Base Pairs Upstream of the c-kit Transcription Initiation Site. *Chem. Eur. J.* **2009,** *15* (46), 12663-12671.
306. Li, W.; Miyoshi, D.; Nakano, S.; Sugimoto, N., Structural competition involving G-quadruplex DNA and its complement. *Biochemistry* **2003,** *42* (40), 11736-44.

307. Kumar, N.; Sahoo, B.; Varun, K. A.; Maiti, S.; Maiti, S., Effect of loop length variation on quadruplex-Watson Crick duplex competition. *Nucleic Acids Res.* **2008,** *36* (13), 4433-42.
308. Dhakal, S.; Yu, Z. B.; Konik, R.; Cui, Y. X.; Koirala, D.; Mao, H. B., G-Quadruplex and i-Motif Are Mutually Exclusive in ILPR Double-Stranded DNA. *Biophys J.* **2012,** *102* (11), 2575-2584.
309. Cui, Y. X.; Kong, D. M.; Ghimire, C.; Xu, C. X.; Mao, H. B., Mutually Exclusive Formation of G-Quadruplex and i-Motif Is a General Phenomenon Governed by Steric Hindrance in Duplex DNA. *Biochemistry* **2016,** *55* (15), 2291-2299.
310. Sutherland, C.; Cui, Y. X.; Mao, H. B.; 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.
311. Brown, R. V.; Wang, T.; Chappeta, V. R.; Wu, G.; Onel, B.; Chawla, R.; Quijada, H.; Camp, S. M.; Chiang, E. T.; Lassiter, Q. R.; Lee, C.; Phanse, S.; Turnidge, M. A.; Zhao, P.; Garcia, J. G. N.; Gokhale, V.; Yang, D.; Hurley, L. H., The Consequences of Overlapping G-Quadruplexes and i-Motifs in the Platelet-Derived Growth Factor Receptor beta Core Promoter Nuclease Hypersensitive Element Can Explain the Unexpected Effects of Mutations and Provide Opportunities for Selective Targeting of Both Structures by Small Molecules To Downregulate Gene Expression. *J. Am. Chem. Soc.* **2017,** *139* (22), 7456-7475.
312. 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.

Chapter 2

Stabilization of i-Motif Structures by 2'-β-Fluorination of DNA



i-Motif structure of 5'-dTCC(fC)(fC)C-3' stabilized by
2'F-araC·2'F-araC$^+$ base pairing
$pH_{1/2}$ = 6.7 ± 0.2, $\Delta pH_{1/2}$ = +0.6

The majority of this chapter is reproduced from: "Stabilization of i-motif structures by 2'-β-fluorination of DNA", **Hala Abou Assi**, Robert W. Harkness V, Nerea Martın-Pintado, Christopher J. Wilds, Ramon Campos-Olivas, Anthony K. Mittermaier, Carlos González, Masad J. Damha, *Nucleic Acids Research*, **2016**, *44*, 4998-5009.

"If you are working on something exciting that you really care about, you don't have to be pushed. The vision pulls you."
–Steve Jobs

2.1 Introduction

Since the discovery of the DNA double helix, the understanding of DNA structure has expanded considerably, with a plethora of polymorphs now known.[1] Two structures have attracted considerable interest due to their possible involvement in telomere maintenance and gene regulation: a) the G-quadruplex (G4),[2,3] which consists of stacked guanine tetrads stabilized by monovalent cations (K^+/Na^+); and b) the i-motif,[4] a cytosine-rich structure consisting of two hemi-protonated ($C{\cdot}C^+$) parallel duplexes intercalated in an antiparallel orientation.[5,6] There is increasing evidence for the occurrence of i-motif forming sequences throughout the human genome such as centromeres,[7,8] telomeres,[9,10] and oncogene promoter regions.[11,12] These structures are also attracting much interest in nanotechnology, for example in the control of macromolecular structure assembly[13] and as pH-responsive switches.[14,15]

The i-motif is a compact DNA structure characterized by short base-pairing distances (3.1 Å), a 12-16° helical twist between adjacent $C{\cdot}C^+$ base pairs (**Figure 2.1A**), and close sugar-sugar contacts stabilized by CH···O interactions.[16] As discussed in the introduction, i-motif structures can have different intercalation topologies known as 3'E and 5'E (**Figure 1.16**). In the 3'E topology, the terminal $C{\cdot}C^+$ base pair is at the 3'-end while the 5'E has the terminal $C{\cdot}C^+$ base pair at the 5'-end.[9] Repulsive interactions exist between the charged C-imino protons, and between the phosphate groups across the crowded narrow groove. *In vitro*, unmodified i-motif structures are stable within a narrow pH range (~3.5-5.5), with maximum stability occurring at pH equal to the pK_a of the cytosine N3. At pHs above the cytosine N3 pK_a (4.4), the stabilizing $C{\cdot}C^+$ base pairing is lost and the i-motif structure is denatured. However, it is likely that *in vivo* factors such as negative superhelicity,[17] cellular proteins,[18] and molecular crowding conditions[19] stabilize i-motif DNA structures at physiological pH. To help compensate for the absence of these factors *in vitro*, small molecule ligands and certain metal ions (Cu^{+2}; Ag^+) have been introduced to promote i-motif folding at neutral pH.[20] Among the sugar and phosphate modifications so far reported, none have provided sufficient stability at neutral pH to permit biochemical studies *in vitro*.[21-25]

The present study was prompted by observations that arabinose (2'-OH),[26] but not ribose (2'-OH or 2'-F),[27,28] is well tolerated within the i-motif structure. We hypothesized that 2'F-arabinose (2'F-araC) would enhance inter-strand CH···O interactions within the narrow grooves

since sequential FC-H2'···O4' interactions have been previously observed in duplexes[29] and G-quadruplex structures.[30] Furthermore, such a modification could potentially stabilize the i-motif by augmenting ion-dipole and stacking interactions of $C \cdot C^+$ base pairs.[16] To explore the effect of 2'-β-fluorination on the structure and stability of different families of i-motif structures, we introduced 2'F-araC modifications in various i-motif forming sequences. We show that replacing 2-deoxyribose sugar with 2-deoxy-2-fluoroarabinose enables the formation of stable i-motif structures over a wide pH range with thermal stabilizations up to 30 °C at neutral pH.

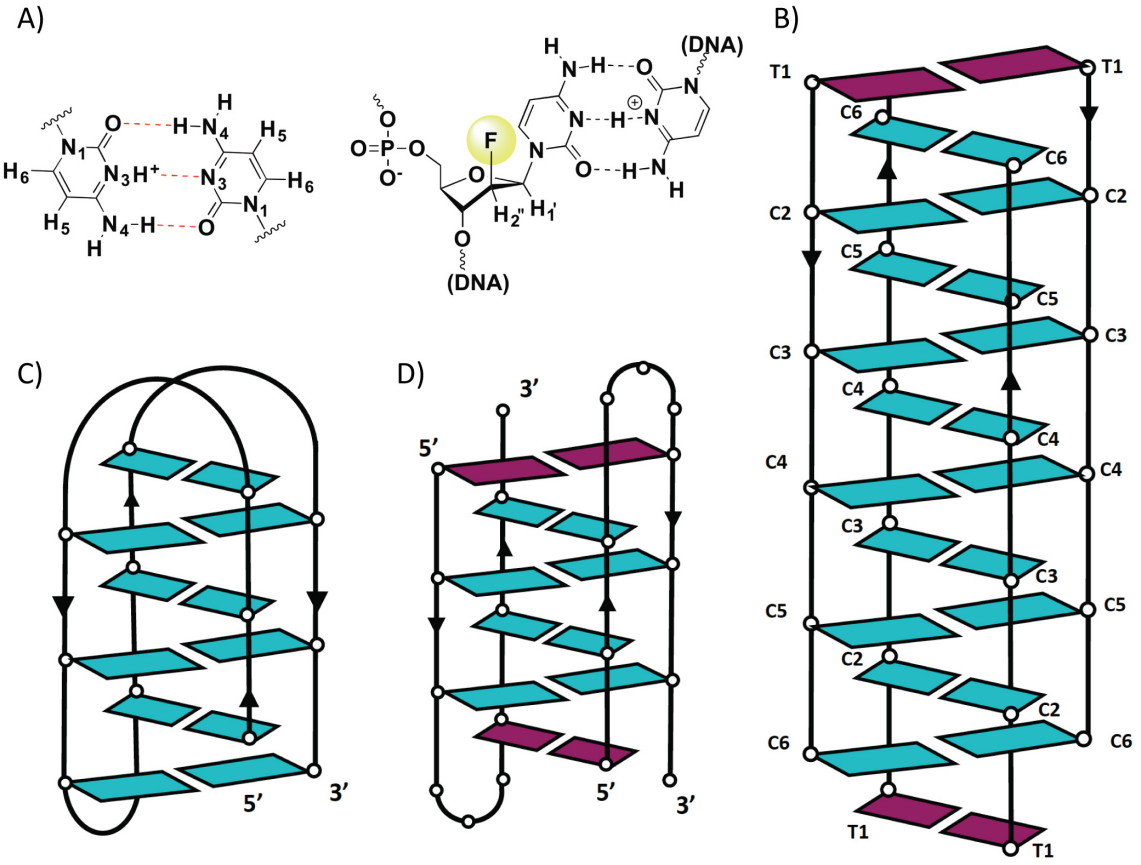


Figure 2.1: A) $C \cdot C^+$ base pair (left) and 2'F-araC$\cdot C^+$ base pair (right). B) Schematic representation of the tetrameric i-motif formed by dTCCCCC (H-1). C) Schematic representation of human telomeric repeat (HT-0). D) Schematic representation of the centromeric A-box (HC-0). Color code: dC are shown in cyan and capping dT are shown in magenta. Residues in the loops are not shown.

2.2 The Effect of 2'F-araC Substitutions on Thermal and pH Stability of i-Motif Structures

Sequences of the oligonucleotides prepared for this study are provided in **Table 2.1**. The first series (named H) is derived from dTCCCCC, a hexanucleotide that associates into a tetrameric i-motif structure (**Figure 2.1B**), whereas the HT and HC series are derived from human telomeric and human centromeric DNA sequences, respectively. Their control unmodified sequences, HT-0 and HC-0, form monomeric and dimeric i-motif structures respectively (**Figures 2.1C** and **2.1D**).

UV melting experiments showed a significant increase in $T_{1/2}$ at acidic and neutral pH in all the monomeric (HT) and dimeric (HC) sequences studied herein (**Figure 2.2** and **Table 2.1**). Here, we use $T_{1/2}$ to refer to the midpoint of the dissociation transition, owing to the observed hysteresis between forward and reverse scans for many of the studied sequences. Although very similar spectral features (NMR and CD) were observed for the native (HT-0) and modified telomeric sequences (e.g. HT-2 and HT-4), the modified structures were significantly more stable at both pH 5.0 and 7.0 relative to the unmodified strands (**Table 2.1**). HC-4 and HT-2 exhibit $T_{1/2}$ values of 28.2 °C and 32.6 °C at pH 7.0, respectively. It is interesting to note that $\Delta T_{1/2}$ values obtained between modified and control oligonucleotides were independent of the temperature gradient used in the melting experiments 0.5 °C/min (**Tables 2.1**) *versus* 0.2 °C/min (**Table AII.1**). In addition, the dissociation/association profiles obtained at 0.5 and 0.2 °C/min for every sequence studied were virtually identical, indicating that the measured processes were the same, despite the presence of hysteresis.

The same stabilization effect was observed in the majority of the tetrameric sequences (**Tables 2.2**). Melting of H-1 was detected at pH 3.5-5.5, whereas melting of the modified sequences was observed over a broader pH range (3.0-7.0). However, the denaturation curves of the tetrameric structures were more complex, showing more than one transition, which suggests the occurrence of several species. The most substituted tetrameric sequences (H-5 and H-6) exhibited lower $T_{1/2}$ values, in contrast to the general stabilization effect observed in all the other sequences. These contradictory results prompted us to study the tetrameric sequences in more detail.

Table 2.1: $T_{1/2}$ (°C) values for the (un)modified sequences.[a]

Code	Sequence (5'-3')	$T_{1/2}$ pH 5.0	$\Delta T_{1/2}$ pH 5.0	$T_{1/2}$ pH 7.0	$\Delta T_{1/2}$ pH 7.0
H-1	dTCCCCC	51.5	-	-	-
H-2	dTCC(fC)CC	10.0, 53.6	+2.1	-	-
H-3	dTC(fC)(fC)CC	25.7, 58.7	+7.2	-	-
H-4	dTCC(fC)(fC)C	21.5, 57.0	+5.5	-	-
H-5 [b]	dTC(fC)(fC)(fC)C	32.0	-	-	-
H-6 [b]	dTf(CCCCC)	37.5	-	10.0	-
HC-0	dTCCTTTTCCA	19.0	-	9.0	-
HC-1	dT(fC)CTTTTCCA	23.0	+4.0	11.0	+2.0
HC-2	dT(fC)(fC)TTTTCCA	31.6	+12.6	24.1	+15.1
HC-3	dT(fC)CTTTT(fC)CA	34.1	+15.1	27.1	+18.1
HC-4	dT(fC)(fC)TTTT(fC)(fC)A	36.2	+17.2	28.2	+19.2
HT-0	$d(CCCTAA)_3C_3$	60.0	-	14.0	-
HT-1	$d(C(fC)CTAA)_3C(fC)C$	67.5	+7.5	23.0	+9.0
HT-2	$d(C(fC)(fC)TAA)_3C(fC)(fC)$	77.1	+17.1	32.6	+18.6
HT-3	$d((fC)(fC)CTAA)_3(fC)(fC)C$	74.1	+14.1	26.6	+12.6
HT-4	$d((fC)(fC)(fC)TAA))_3(fC)_3$	80.2	+20.2	31.2	+17.2

[a]Oligonucleotide concentration was 4.6 µM in single strands for the hexamer sequences and 4.0 µM for HC and HT. $T_{1/2}$ data were calculated from UV-Visible spectroscopy thermal denaturation profiles where $T_{1/2}$ corresponds to the midpoint of the dissociation transition obtained at 0.5 °C/min. $\Delta T_{1/2}$ values were calculated for the main melting transitions relative to the respective unmodified strands. (-) indicates sequences for which no melting transition was detected, or an absence of $\Delta T_{1/2}$.

[b]Sequences H-5 and H-6 exhibited lower $T_{1/2}$ values compared to the control (H-1). This is due to the melting of a dimeric structure rather than an i-motif structure. Therefore $\Delta T_{1/2}$ cannot be calculated since the comparison with H-1 would be invalid.

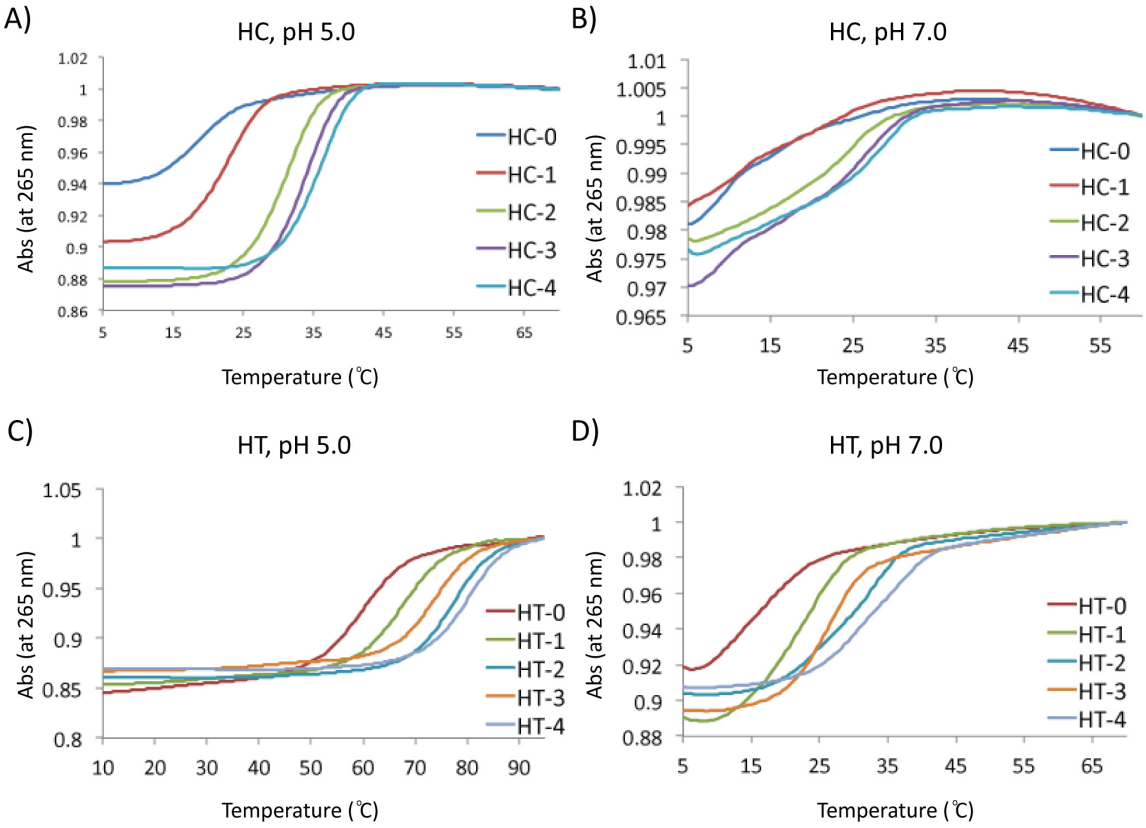


Figure 2.2: $T_{1/2}$ spectra recorded at 265 nm and 0.5 °C/min in 10 mM NaP_i buffer and 4 µM strand concentration. A) HC i-motifs at pH 5.0. B) HC i-motifs at pH 7.0. C) HT i-motifs at pH 5.0. D) HT i-motifs at pH 7.0.

Table 2.2: $T_{1/2}$ (°C) values for the hexamer sequences at different pH values.

pH	H-1		H-2		H-3		H-4		H-5		H-6	
	$T_{1/2}$	%H	$T_{1/2}$	%H	$T_{1/2}$	%H	$T_{1/2}$	%H	$T_{1/2}$	%H	$T_{1/2}$	%H
3.0	-	-	-	-	32.7	9.9	35.5	8.6	13.5	12.2	31.7	20.6
3.5	36.5	8.7	39.1	19.0	12.7 46.7	1.6 15.5	10.5 47.0	1.0 11.3	30.5 51.5	14.7 3.7	44.6	19.8
4.0	47.0	12.5	50.1	15.8	25.2 57.2	9.8 6.1	24.0 56.0	8.3 6.6	39.0	14.3	48.7	15.7
5.0	51.5	23.3	10.0 53.6	1.3 16.9	25.7 58.7	12.0 5.3	21.5 57.0	9.3 5.8	32.0	13.0	37.5	13.1
5.0 *	9.5 53.0	2.7 6.6	17.1 55.1	9.1 3.6	24.7 59.7	13.5 1.1	24.0 58.5	12.4 1.3	35.0	12.4	41.6	11.8
5.5	8.0 51.0	1.1 6.4	9.6 53	2.9 7.8	16.2 57.2	9.5 0.4	18.6 56.5	11.2 0.7	28.0	11.6	32.2	12.1
6.0	-	-	7.1 45.6	0.5 4.9	10.2 50.2	3.0 3.7	11.0 49.5	3.8 0.8	19.0	11.1	21.0	11.9
6.0 *	-	-	10 .1 49.1	3.5 2.0	11.7 52.7	6.3 1.5	11.5 52.5	5.4 1.8	24.0	11.6	26.2	11.4
6.5	-	-	-	-	8.2	3.0	10.0	2.8	12.5	7.7	18.2	10.1
7.0	-	-	-	-	-	-	-		-	-	10.0	4.0

Oligonucleotide concentration was 4.6 μM in single strands. $T_{1/2}$ were recorded at 265 nm. Buffers used were 10 mM sodium citrate (pH 3.0 – 4.0) and 10 mM sodium phosphate (pH 5.0 – 7.0). pH 5.0* (10 mM sodium acetate buffer). pH 6.0* (10 mM sodium cacodylate buffer). (-) not detected. Melting curves were recorded at a heating rate of 0.5 °C/min. %H corresponds to hyperchromicity.

Circular dichroism provides a convenient way to detect the formation of i-motifs since their spectra exhibit characteristic negative and positive bands at around 265 and 285 nm, respectively.[31] As previously reported, the model sequence dTCCCCC (H-1) exhibited a positive band that decreases in magnitude and becomes blue shifted as the pH is raised above 5.5 (**Figure 2.3A**). By contrast, some of the modified hexanucleotides (e.g., H-3, H-4, H-5, and H-6) exhibited this characteristic i-motif CD signature at pH 6.5 (**Figures 2.3B** and **AII.1**). Similarly, the modified centromeric and telomeric sequences maintained the i-motif CD signature at neutral pH (**Figures 2.3C** and **2.3D**).

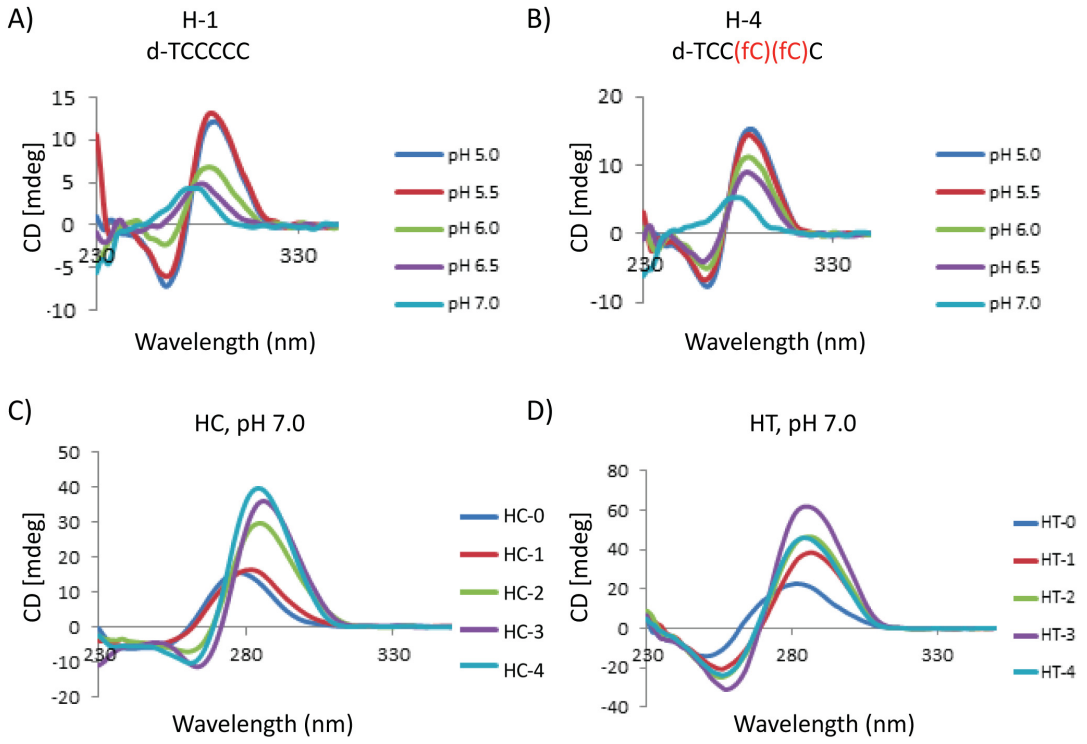


Figure 2.3: A) and B) pH-dependent CD spectra at 30 µM and T = 5 °C for H-1 and H-4, respectively. C) CD spectra for centromeric sequences at pH 7.0, 100 µM, and T = 5 °C. D) CD spectra for telomeric sequences at pH 7.0, 50 µM, and T = 5 °C.

pH titration experiments were performed for the modified and unmodified i-motif structures to determine the pH at which 50% of the strands are folded into i-motif structures ($pH_{1/2}$). The titration profiles were fit to a standard titration model assuming a single protonation event in order to extract populations of folded (protonated) and unfolded (deprotonated) states as a function of pH (**Figure 2.4**). The "fraction protonated" profiles indicate that the modified i-motif structures were markedly stabilized at pH 7.0 relative to the corresponding unmodified references. For example, H-4 and H-6 were roughly 40% and 50% folded at pH 7.0 respectively, whereas H-1 was 10% folded.

The pH stabilizations translate to the centromeric and telomeric sequences where the fraction protonated profiles for the modified HC-3 and HT-4 sequences exhibited folded populations of approximately 60% and 90% at pH 7.0 respectively, compared to the unmodified HC-0 and HT-0 which were roughly 10% and 30% folded (**Figure 2.4**). This 6- and 3-fold increase in folded population is translated to an increase in thermal stability for HC-3 ($\Delta T_{1/2}$ = +18.1 °C) and HT-4 ($\Delta T_{1/2}$ = +17.2 °C).

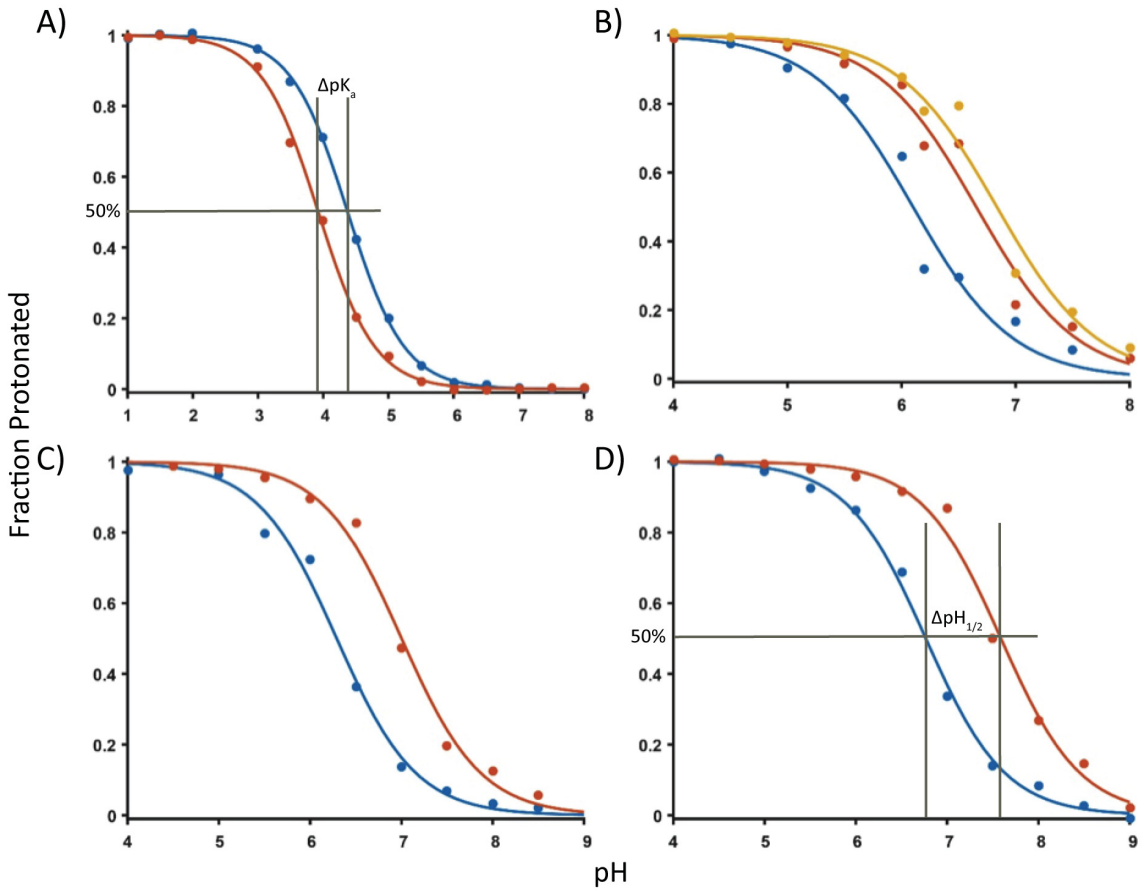


Figure 2.4: Nucleoside pK_a and i-motif $pH_{1/2}$ data. A) Nucleoside titration curves: dC data are in blue, 2'F-araC data are in red. UV data were collected at 280 nm. B) Tetramer titration curves: H-1 are in blue, H-4 are in red, and H-6 are in yellow. C) Centromeric titration curves: HC-0 are in blue and HC-3 are in red. D) Telomeric titration curves: HT-0 are in blue and HT-4 are in red. Tetrameric, centromeric, and telomeric CD data as a function of pH were collected at 284 nm at T = 5 °C. Experimental data are shown as colored filled circles, fits are shown as colored lines. Experimental data were fit to $CD_{obs} = CD_{high} + (CD_{low} - CD_{high})/(1 + 10^{pH-pKa})$.

The $pH_{1/2}$ differences from the modifications that contributed to the large increase in thermal stability are listed in **Table 2.3**. The $pH_{1/2}$ data of the i-motif structures exhibited an opposite trend compared to the nucleoside pK_as (2'F-araC and dC were 3.9 and 4.4, respectively). This indicates that the effective pK_a of 2'F-araC residues is strongly affected by the three-dimensional structure of the i-motif. Most likely, this large pK_a shift is due to the hydrogen-bonding network and electrostatic interactions in the neighborhood of the protonation sites.[32]

Table 2.3: Nucleoside pK_a[a] and i-motif $pH_{1/2}$ values.[b]

Code	Nucleoside	pK_a	ΔpK_a
dC	2'-deoxycytidine	4.4 ± 0.02	-
2'F-araC	2'-deoxy-2'-fluoroarabinocytidine	3.9 ± 0.05	-0.5
Code	**Sequence**	**$pH_{1/2}$**	**$\Delta pH_{1/2}$**
H-1	5'-dTCCCCC-3'	6.1 ± 0.2	-
H-4	5'-dTCC(fC)(fC)C-3'	6.7 ± 0.2	0.6
H-6	5'-dTf(CCCCC)-3'	6.8 ± 0.3	0.7
HC-0	5'-dTCCTTTTCCA-3'	6.3 ± 0.2	-
HC-3	5'-dT(fC)CTTTT(fC)CA-3'	7.0 ± 0.2	0.7
HT-0	5'-d(CCCTAA)$_3$CCC-3'	6.8 ± 0.1	-
HT-4	5'-d((fC)(fC)(fC)TAA)$_3$(fC)(fC)(fC)-3'	7.6 ± 0.3	0.8

[a]The pK_a values for dC and 2'F-araC were calculated from the plot of UV absorbance at 280 nm in 10 mM NaP_i buffer as a function of pH. The $pH_{1/2}$ values for i-motif structures were calculated from the plot of molar ellipticity at 284 nm *versus* pH in 10 mM NaP_i buffer at 5 °C. [b]CD versus pH data were fit to a standard titration model involving a single protonation event using $CD_{obs} = CD_{high} + (CD_{low} - CD_{high})/(1 + 10^{pH-pKa})$. Errors were calculated according to the variance-covariance method.[33]

2.3 Studying i-Motif Formation by ^{1}H-NMR

i-Motif formation can be also detected by ^{1}H-NMR. All the NMR data reported in this chapter were obtained in collaboration with Prof. Carlos González at the CSIC in Madrid. Signals at 15-16 ppm, characteristic of cytosine imino protons in hemi-protonated $C{\cdot}C^+$ base pairs, were observed for most of the modified sequences over a wide temperature and pH range (**Figures 2.5-2.7 and AII.2**). For instance, the control i-motif (H-1) exhibited no imino signals at pH 7.0 (**Figure 2.5A**) whereas i-motif structures formed by H-3, H-4, H-5, and H-6 maintained $C{\cdot}C^+$ base pairing at neutral pH and at relatively high temperatures (35 °C) (**Figures 2.5B** and **AII.2**). This effect was more pronounced in the case of the dimer centromeric sequences (**Figure 2.6**) and monomer human telomeric sequences (**Figure 2.7**), in which ^{1}H-NMR spectra exhibited imino signals at 15-16 ppm, which were clearly visible at physiological temperatures and neutral pH.

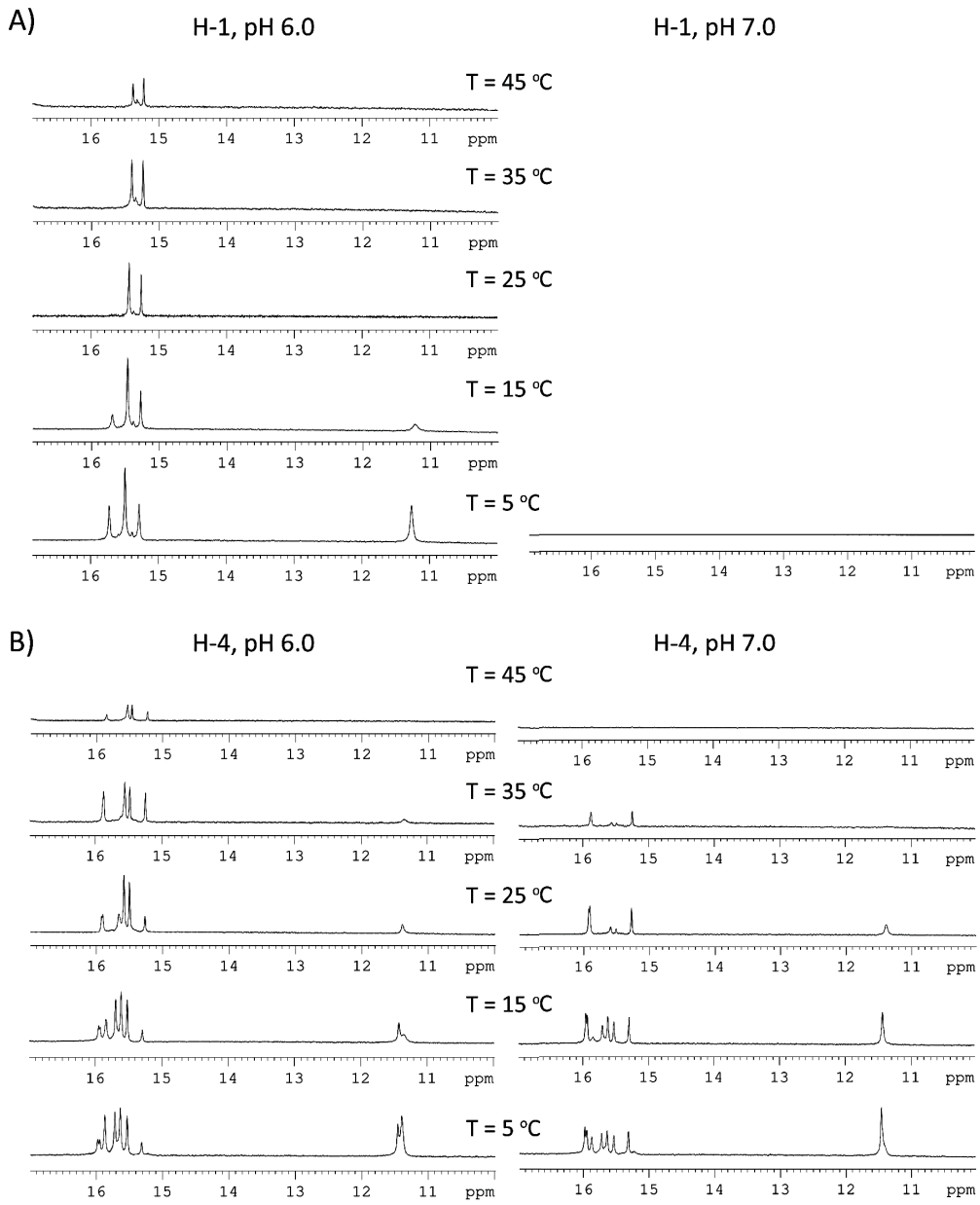


Figure 2.5: A) ^{1}H-NMR melting experiments for H-1 (5'-TCCCCC-3', 2.0 mM). At pH 7.0 and T = 5 °C, no imino peaks were observed. b) ^{1}H-NMR melting experiments for H-4 (5'-TCC(fC)(fC)C-3', 2.53 mM).

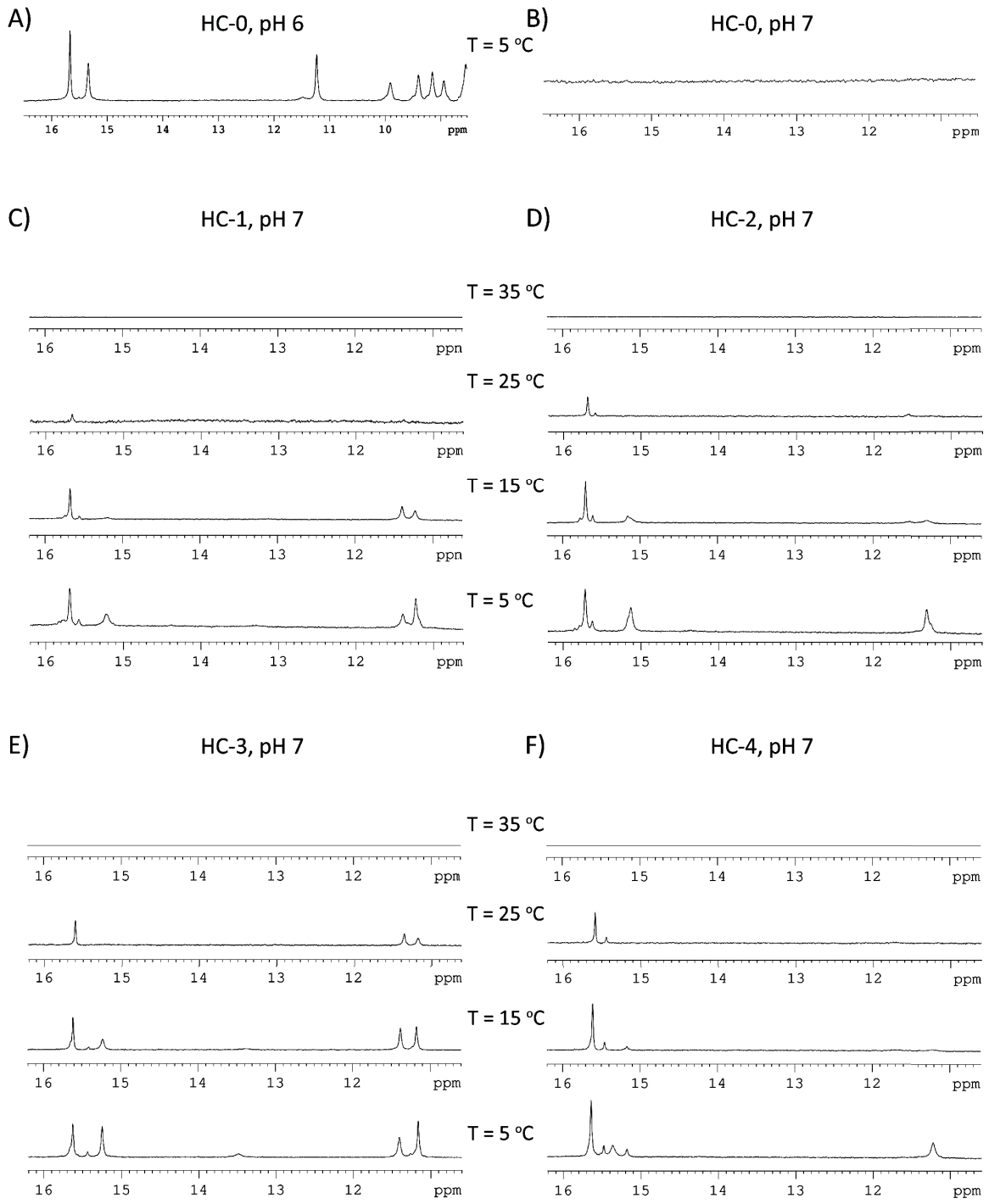


Figure 2.6: ^{1}H-NMR melting experiments (5 °C to 35 °C) for centromeric sequences at pH 7.0. At neutral conditions, no imino peaks were observed at 5 °C for HC-0. However, in the modified centromeric sequences the imino signals were clearly visible at 25 °C.

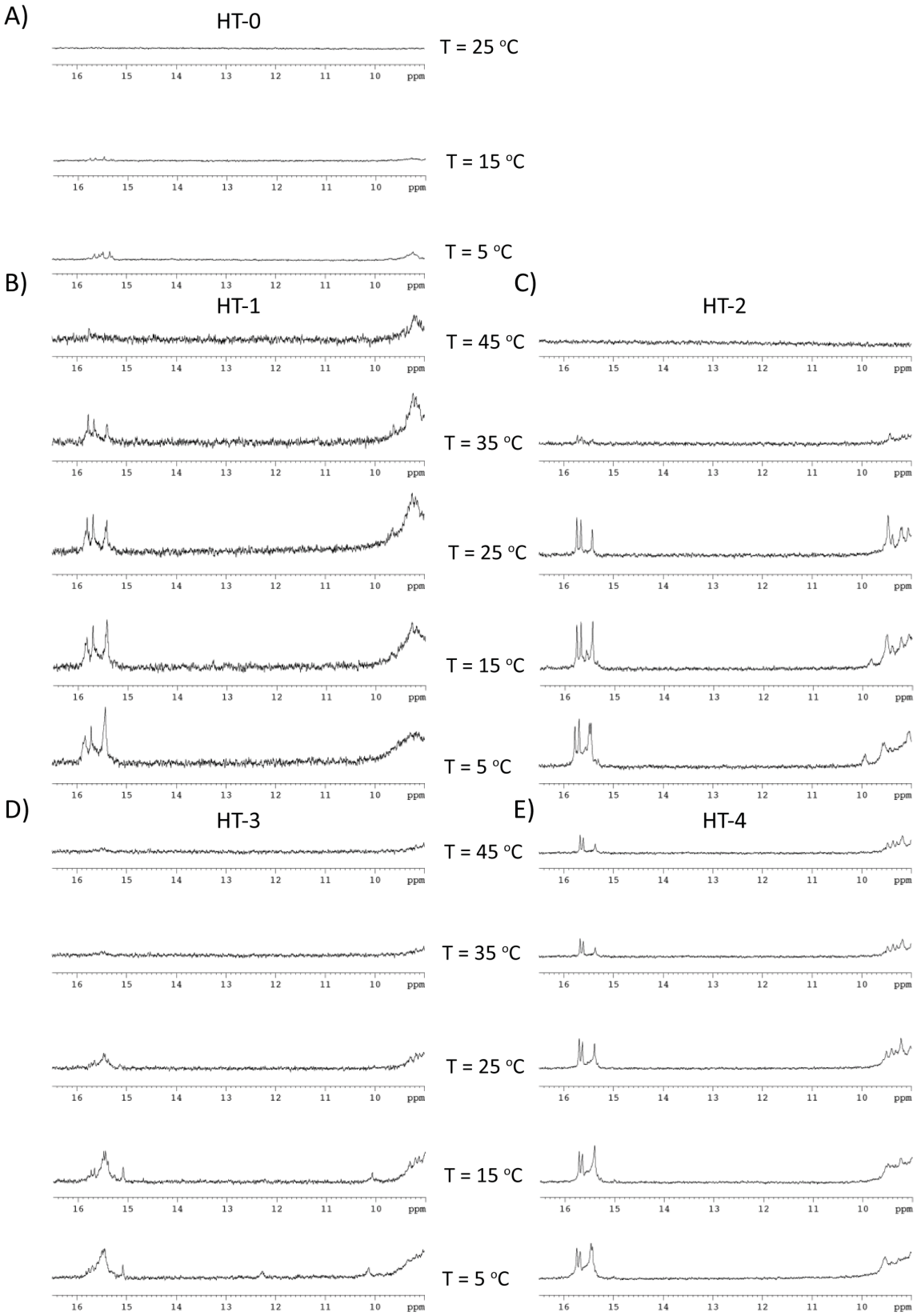


Figure 2.7: ^{1}H-NMR melting experiments (5 °C to 45 °C) for telomeric sequences at pH 7.0.

2.4 Determining the Molecularity of the i-Motif Structures *via* Gel Electrophoretic Experiments and ^{19}F-NMR

Native PAGE experiments confirmed the formation of more than one species in the tetrameric i-motif structures at low temperature (5 °C, **Figures 2.8** and **2.9**). For example, the denaturation curve of the singly modified oligomer (H-2, 4.6 μM; pH 5.0) exhibited a minor transition with a $T_{1/2}$ of 10.0 °C and a major transition with a $T_{1/2}$ of 53.6 °C that corresponded closely to that of the control sequence ($T_{1/2}$ of 51.5 °C). This was also evident by native PAGE, where H-2 (40 μM, **Figure 2.8**, lane 4) appeared as a mixture of closely moving bands that migrated with the same mobility as the tetrameric control structure H-1.

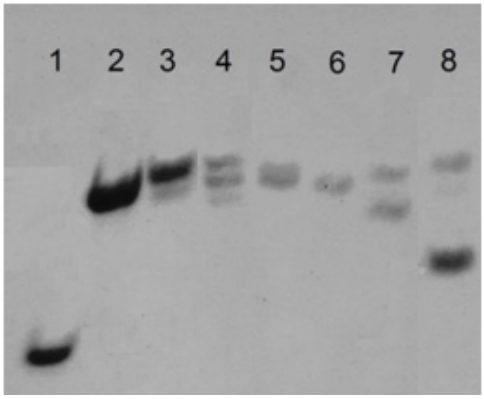


Figure 2.8: Native gel, pH 5.0, 40 μM strand concentration. Lane 1: dT_{12}, Lane 2: dT_{24}, Lane 3: (H-1) TCCCCC, Lane 4: (H-2) TCC(fC)CC, Lane 5: (H-3) TC(fC)(fC)CC, Lane 6: (H-5) TC(fC)(fC)(fC)C, Lane 7: (H-4) TCC(fC)(fC)C, Lane 8: (H-6) Tf(CCCCC).

Results obtained for H-2, H-3, H-4 and H-5 from concentration-dependent gel electrophoresis experiments were consistent with the formation of topologically similar i-motif structures at all concentrations studied (4.6-200 μM; **Figure 2.9**). A major species was observed for the unmodified control sequence at 4.6-20 μM, with a minor component appearing on the gel at >40 μM strand concentration (**Figure 2.9**).[4] Interestingly, the fully substituted sequence (H-6) exhibited an aberrant gel electrophoretic mobility (**Figure 2.9**), a significantly lower $T_{1/2}$ value (37.5 °C, **Table 2.1**), and fast association-dissociation kinetics (**Figure 2.10**), all consistent with the formation of a dimeric structure in this case. For instance, at pH 5.0 and 40 μM, H-6 appeared as two bands: one fast moving band and one band, which co-migrated with the H-1 control. This slow moving band corresponded to a tetrameric structure, which was the predominant species at higher strand concentration (100-200 μM).

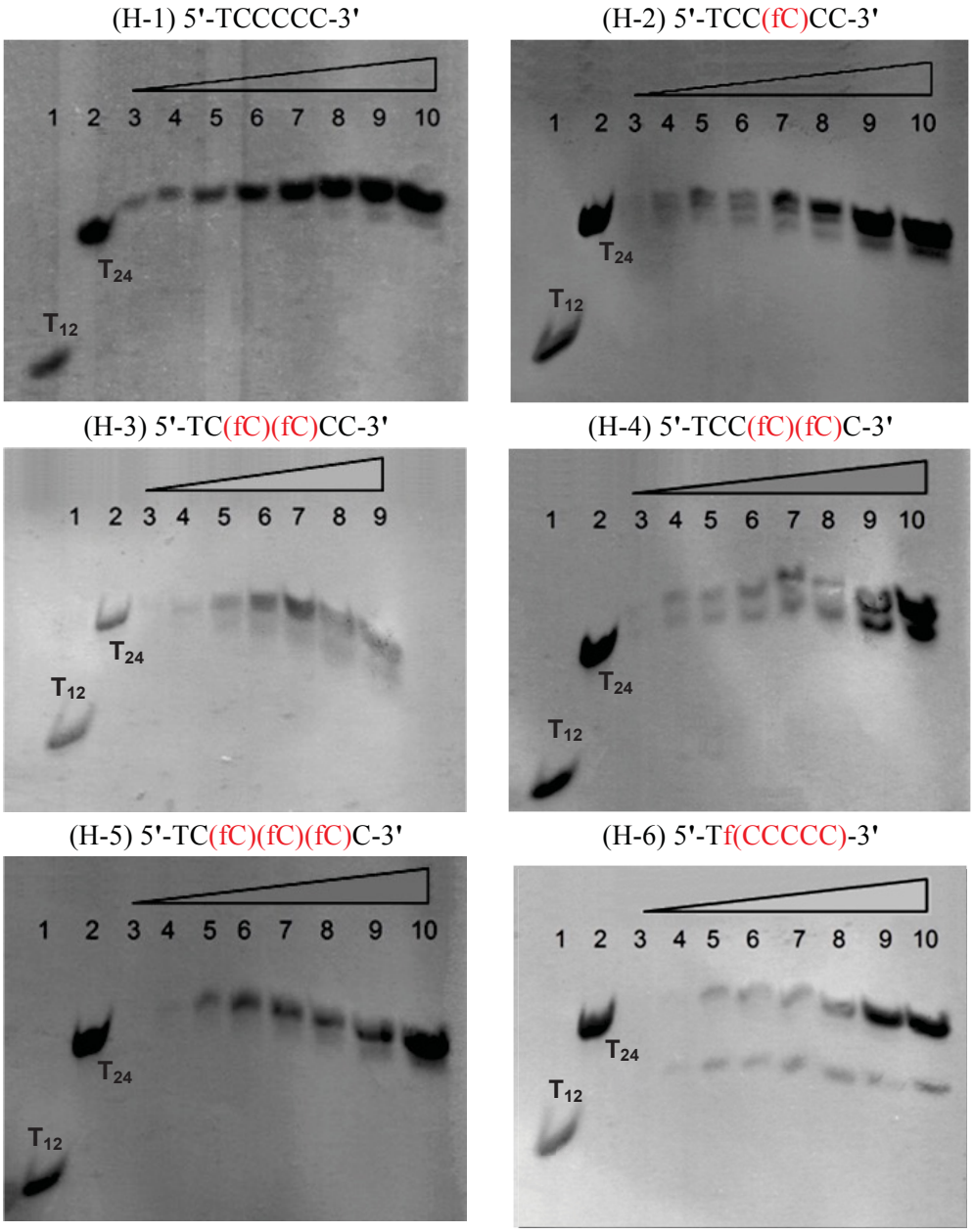


Figure 2.9: Concentration-dependent native gels, pH 5.0. Lane 1: dT_{12}, Lane 2: dT_{24}, Lane 3: 4.6 µM, Lane 4: 10 µM, Lane 5: 20 µM, Lane 6: 40 µM, Lane 7: 80 µM, Lane 8: 100 µM, Lane 9: 140 µM, Lane 10: 200 µM.

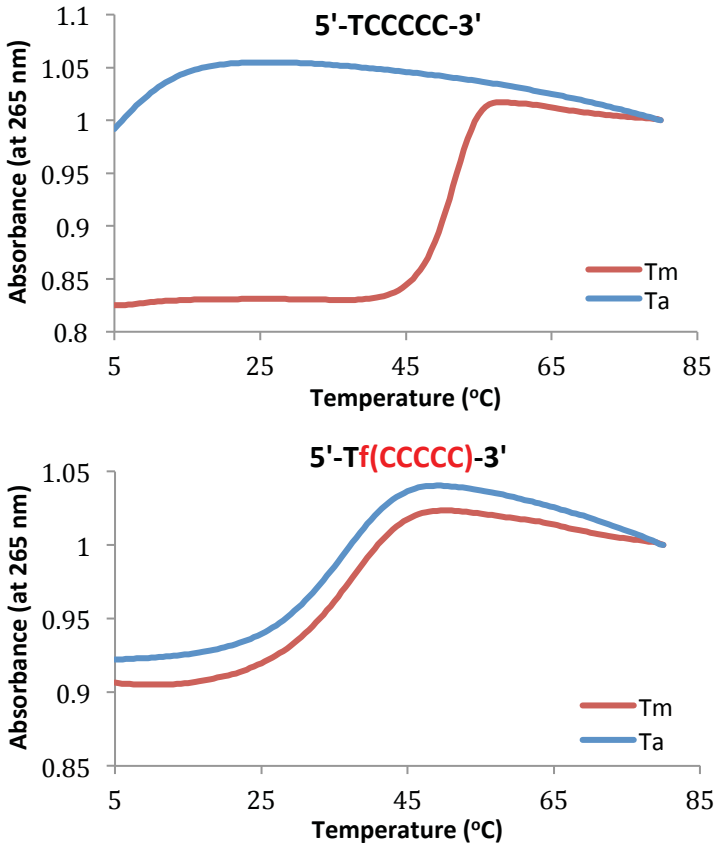


Figure 2.10: Denaturation profiles of H-1 (top) and H-6 (bottom) obtained at 265 nm and a heating rate of 0.5 °C/min. Strand concentration was 4.6 μM in 10 mM NaP_i buffer, pH = 5.0. T_m corresponds to the melting curve (5 - 85 °C) and T_a corresponds to the annealing/cooling curve (85 - 5 °C).

The wide dispersion of fluorine chemical shifts in the ^{19}F-NMR spectra facilitated the study of multiple species in equilibrium under various experimental conditions.[34] In most of the tetrameric structures, the number of ^{19}F signals was not consistent with a single conformation, in agreement with PAGE data. ^{19}F resonances from different species, including the unfolded oligonucleotide, were observed simultaneously in the whole range of temperatures explored. This indicated that the equilibria were slow on the NMR timescale at all temperatures, as observed previously with other 2'F-araC modified oligonucleotides.[29,35] With the exception of sequences H-5 and H-6, ^{19}F spectra recorded at high (1-2 mM) and low (0.1 mM) concentrations were very similar (**Figures 2.11** and **AII.3**).

As observed in UV melting and gel electrophoresis experiments, the fully substituted sequence (H-6) exhibited distinctive features. ^{19}F-NMR spectra for H-6 at high and low

oligonucleotide concentration were completely different (**Figure 2.11**). Thus, ^{19}F-NMR spectra confirmed the presence of species of different molecularity, as suggested by UV and gel electrophoretic mobility experiments. ^{19}F signals of the low concentration species were observed in the high concentration sample prepared under a fast annealing procedure. On the other hand, the high concentration spectrum was partially recovered after storing the low concentration sample for several months at T = 5 °C (**Figure 2.12**).

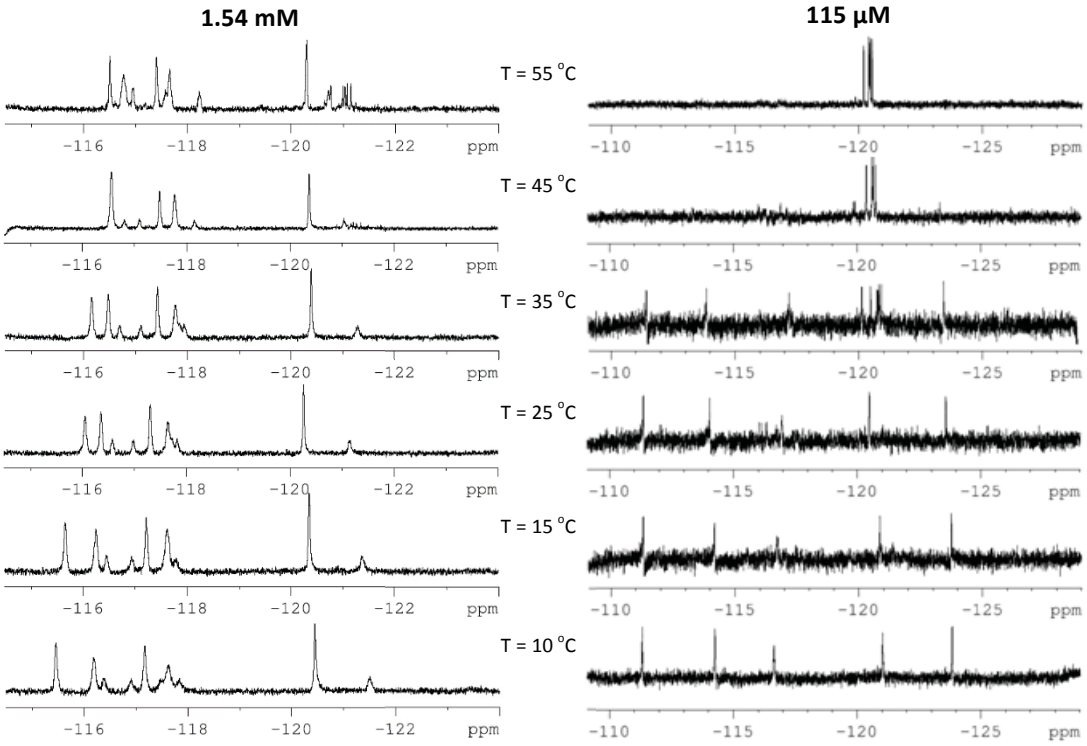


Figure 2.11: ^{19}F-NMR melting experiments for H-6 5'-Tf(CCCCC)-3', pH 5.0.

These observations indicate that the dimeric species is kinetically favored. Such strong kinetic bias explains the apparent destabilization observed in H-6, since the $T_{1/2}$ determined by UV (37.5 °C) corresponds to a kinetically trapped species detected by NMR at low concentrations. Interestingly, ^{1}H-NMR spectra of H-6 recorded in conditions in which the dimeric species was predominant exhibited signals in the 15-16 ppm region, as expected for hemiprotonated cytosines (**Figure 2.13**). Most probably, this structure is a parallel homo-duplex stabilized by C·C^{+} base pairs.[36,37] Since there was no indication of the formation of this dimeric species in the unmodified oligonucleotide, we must consider that 2'F-araC substitutions may also influence the stabilization of parallel duplexes, which in this case can be considered as

intermediates in the formation of i-motif structures. ^{19}F spectra suggested that a similar effect may also occur in H-5. Therefore, the only two exceptions in **Table 2.1** can be explained by the formation of kinetically trapped dimeric structures. To extract more reliable thermodynamic data on the tetrameric i-motif formation, we made use of calorimetric techniques such as differential scanning calorimetry.

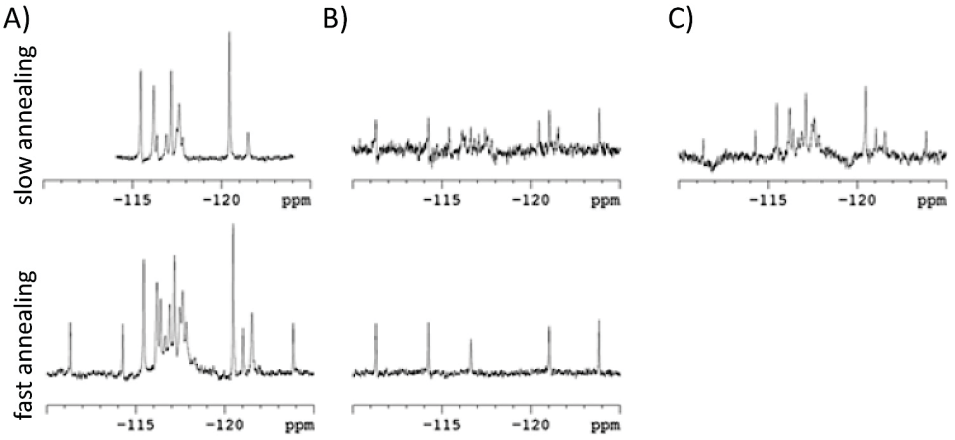


Figure 2.12: ^{19}F-NMR recorded at pH 5.0 and 5 °C for H-6 5'-Tf(CCCCC)-3' after slow and fast annealing procedures. A) Spectra for the high concentration sample (1.54 mM). B) Spectra for the low concentration sample (115 µM). C) Spectra for the low concentration sample (115 µM) recorded after several months. Slow annealing procedure: samples were heated to 90 °C for 15 minutes, then cooled slowly to room temperature and stored at 5 °C for at least 16 h before recording the spectra. Fast annealing procedure: samples were heated in the NMR tube at 65 °C for around 15 minutes, with ^{19}F-NMR spectra subsequently recorded at 5 °C.

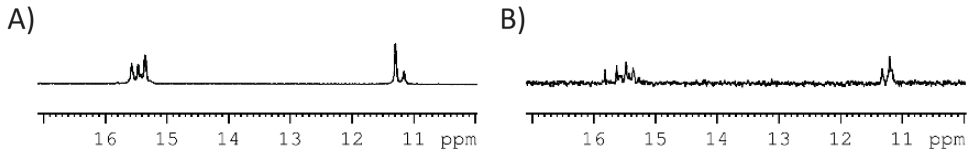


Figure 2.13: ^{1}H-NMR spectra for H-6 5'-Tf(CCCCC)-3' at pH 5.0 and T = 5 °C for A) concentrated sample (1.54 mM) and B) diluted sample (115 µM).

2.5 Thermodynamic Analysis by Differential Scanning Calorimetry

In collaboration with Robert Harkness in the laboratory of Prof. Anthony K. Mittermaier in the Department of Chemistry at McGill University, differential scanning calorimetry (DSC) experiments were performed. DSC is well-suited to detecting the presence of multiple structural populations. We performed DSC experiments on samples H-1 through H-6, finding the main melting transition to occur at higher temperatures with increasing numbers of 2'F-araC

modifications (**Figure 2.14**). With the ~33-fold higher concentrations relative to UV-Vis experiments, the association kinetics were much faster, reflected in the minimal hysteresis across the sample set (**Figure AII.4**). Interestingly, an additional low temperature transition became more prominent as the number of modifications was increased (H-3 through H-6). The DSC profile for H-6 was in greatest contrast, with a wide low temperature shoulder and the highest main transition temperature of the entire set. Positive unfolding ΔC_ps ≈ 2-6 kJ mol^{-1} K^{-1} were found for each structural ensemble.

The DSC profiles were fitted assuming a non-two-state folding behavior, finding in all cases that the extracted structural populations were in close agreement with PAGE and NMR data. H-1 to H-5 were fit with a thermodynamic model assuming the presence of two tetrameric structures, while the model for H6 assumed a tetrameric and dimeric structure, in accordance with gel and NMR results. Thermodynamic parameters resulting from the DSC fits are displayed in **Table 2.4**. Because certain samples exhibited slight hysteresis (**Figure AII.4**), we define $T_{1/2}$ here as the temperature at which respective structural populations were equal to the population of the monomer. For the samples with no hysteresis, this value is the equilibrium T_m. Optimal data quality with minimal hysteresis was found with scan rates of 0.5 °C/min; therefore, we applied this rate across the H-1 to H-6 dataset. The populations extracted from the fit for H-1 show a small fraction of a second tetrameric structure at low temperature. DSC populations for H-2 to H-5 indicate the presence of minor tetrameric structures at low temperature, which became the dominant populations approaching $T_{1/2}$.

These unfolding intermediates may be different topologies or frayed structures, such as those observed in G-quadruplexes and homoduplexes.[38,39] The NMR for H-4 (discussed in the NMR section) suggested that the stable structure at pH 7.0 has opened terminal bases. This might be the case for the unfolding intermediates in H-2 through H-5. H-6 populations indicate the tetrameric structure was favored at low temperature and the dimeric intermediate structure was strongly populated towards the $T_{1/2}$ to give the main transition. Importantly, the population of intermediate states in H-2 to H-6 was consistent with the NMR data.[40,41]

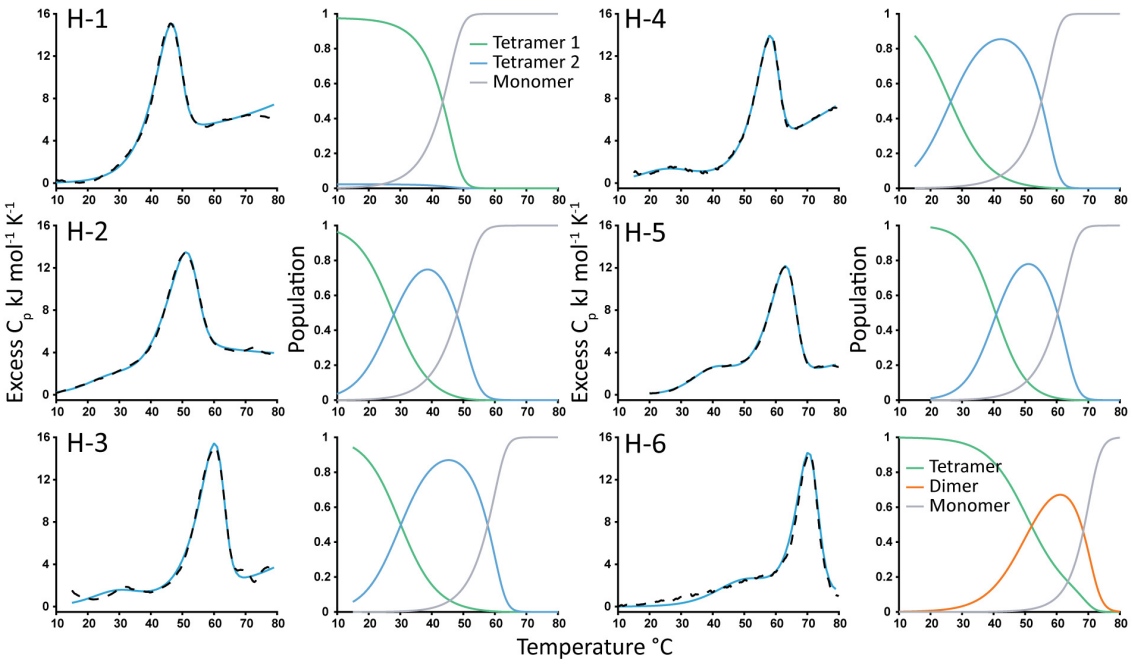


Figure 2.14: Excess C_p profiles for the control and 2'F-araC modified tetrameric i-motif structures. Sequence numbers are indicated in the top left corner of each set. Experimental and fitted excess heat capacities (left panels) are shown as black dashed and blue solid lines, respectively. Structural populations from the fits for each sequence are shown in the panels immediately to the right. Tetramer, dimer, and monomer populations are shown as green and blue, orange, and grey lines respectively.

In addition to identifying structural populations and quantifying their unfolding thermodynamics, the DSC fitting allowed determination of $T_{1/2}$s for individual populations, as well as the global $T_{1/2}$ at which 50% of the structural ensemble was denatured. To quantify the increase in thermal stability due to the modifications, we compared the increase in $T_{1/2}$ for each modified structure relative to the corresponding unmodified structure ($\Delta T_{1/2} = T_{1/2}^{modified} - T_{1/2}^{unmodified}$) with the number of modifications. The relation between $\Delta T_{1/2}$ and number of modifications was strongly correlated (R = 0.98, 0.98, 0.98), showing the $T_{1/2}$s for structures 1, 2, and the global $T_{1/2}$ are increased by 7.5, 5.1, and 5.2 °C respectively with each 2'F-araC modification (**Figure 2.15**).

Table 2.4: Thermodynamic parameters extracted from fits of the tetrameric i-motif DSC melting experiments.

	$-\Delta H_1$[a]	$-\Delta H_2$[a]	$-\Delta S_1$[b]	$-\Delta S_2$[b]	$-\Delta G_1$[a]	$-\Delta G_2$[a]	$T_{1/2(1)}$[c]	$T_{1/2(2)}$[c]	Global $T_{1/2}$[d]
H-1	572.6 ±0.2	572.4 ±3.8	1615 ±35	1583 ±18	61.5 ±0.7	71.4 ±0.7	23.9 ±0.8	43.4 ±0.7	43.5 ±0.2
H-2	692.4 ±4.4	562.3 ±2.7	1960 ±15	1528 ±9	72.1 ±0.4	78.7 ±0.3	38.7 ±0.6	48.0 ±0.7	48.1 ±0.2
H-3	771.8 ±0.7	638.1 ±0.8	2146 ±5	1705 ±2	92.6 ±1.5	98.5 ±0.9	45.7 ±1.2	57.7 ±0.8	57.8 ±1.7
H-4	545.3 ±0.4	422.9 ±0.3	1457 ±10	1049 ±1	84.2 ±0.7	90.9 ±0.3	42.2 ±0.9	55.1 ±0.4	55.2 ±0.3
H-5	663 ±0.4	493.9 ±0.4	1783 ±54	1244 ±37	98.7 ±0.6	100.2 ±0.4	50.6 ±0.4	60.3 ±0.5	60.4 ±0.3
H-6	677.7 ±6.1	205.5 ±4.4	1743.0 ±19	498 ±16	126.0 ±0.4	47.9 ±0.9	63.5 ±1.0	68.4 ±1.2	68.9 ±0.2

[a]ΔH and ΔG expressed in kJ mol^{-1}. [b]ΔS expressed in J mol^{-1} K^{-1}. Thermodynamic parameters were calculated at the global $T_{1/2}$ for the control sequence H-1.[c]$T_{1/2}$s (expressed in ^{o}C) for structures 1 and 2 were defined as the temperatures where the respective populations were equal to the population of the monomer. [d]Global $T_{1/2}$s were defined as the temperatures at which the monomer was 50% populated. Numbered subscripts refer to thermodynamic parameters for melting tetramer 1 and tetramer 2 in sequences H-1 to H-5 and tetramer and dimer respectively in sequence H-6. Errors were calculated according to the variance-covariance method.[33]

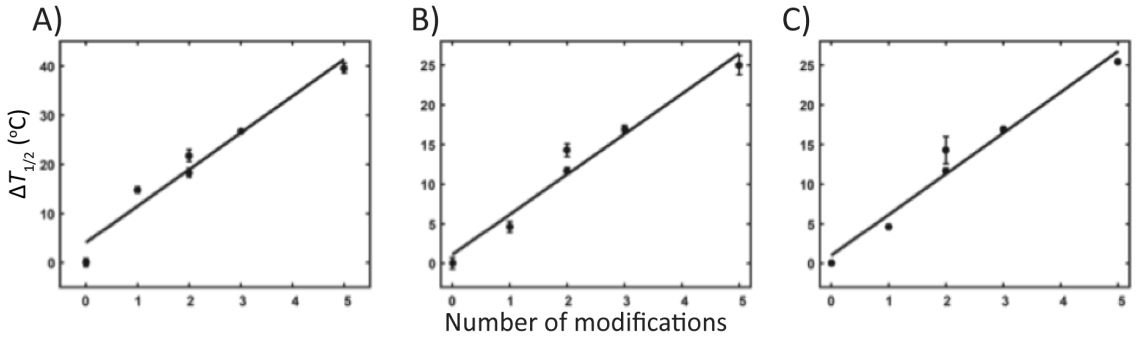


Figure 2.15: $\Delta T_{1/2}$ (°C) as a function of the number of 2'F-araC modifications. A) Modifications increased the stability of structure 1 by 7.5 °C (R = 0.98). B) Modifications increased the stability of structure 2 by 5.1 °C (R = 0.98). C) Modifications increased the global $T_{1/2}$ of the structural ensemble by 5.2 °C (R = 0.98).

2.6 Structural Basis of 2'F-araC Stabilization

To obtain further insight into the origin of the stabilization due to fluoroarabinose modifications, an in-depth structural study was undertaken using NMR spectroscopy. Complete assignment of the NMR spectra of highly repetitive sequences such as those studied here is a difficult task, which is further exacerbated here by the occurrence of multiple folded species in the case of the tetrameric structures. However, the presence of 2'F-araCs in the sequence improved chemical shift dispersion and greatly facilitated 2D NMR analysis. In the case of H-4, we accomplished a complete sequential assignment of the major species at pH 5.0 (**Figure 2.16)**, consistent with a 3'E intercalation topology.

Strong H1'-H1' cross-peaks and amino-H2'/2" NOEs clearly confirmed the formation of an i-motif structure. The stacking order could be determined by following H1'-H1' connectivities along the minor groove between C2-fC5, C2-C6, C6-T1, C3-fC4, and C3-fC5. This pattern was confirmed by the amino-H2'/H2" contacts observed along the major groove between T1-C6, C2-fC5 and C3-fC4 (**Figure 2.16**). The exchangeable proton spectra exhibited several signals between 15-16 ppm, indicative of $C \cdot C^+$ base pairs. These base pairs must have occurred between magnetically equivalent cytosines, since each of these signals exhibited cross-peaks with only two amino protons (8-10 ppm). NMR assignments are specified in **Table AII.2**.

DQF-COSY experiments indicated that deoxyriboses adopted a north conformation (except C6) due to their intense COSY cross peaks between H1'-H2", while 2'F-araCs adopted a south conformation (**Figure 2.17**). The three dimensional structure of H-4 was calculated on the basis of 284 experimental distance constraints using restrained molecular dynamics methods, and following standard procedures previously described by our group.[42] All residues were well defined, with an RMSD of 0.9 Å (**Table AII.3**). The final AMBER energies and NOE terms were reasonably low in all the structures, with no distance constraint violation >0.3 Å. The coordinates of the final refined structures were deposited in the PDB (code:2N89).

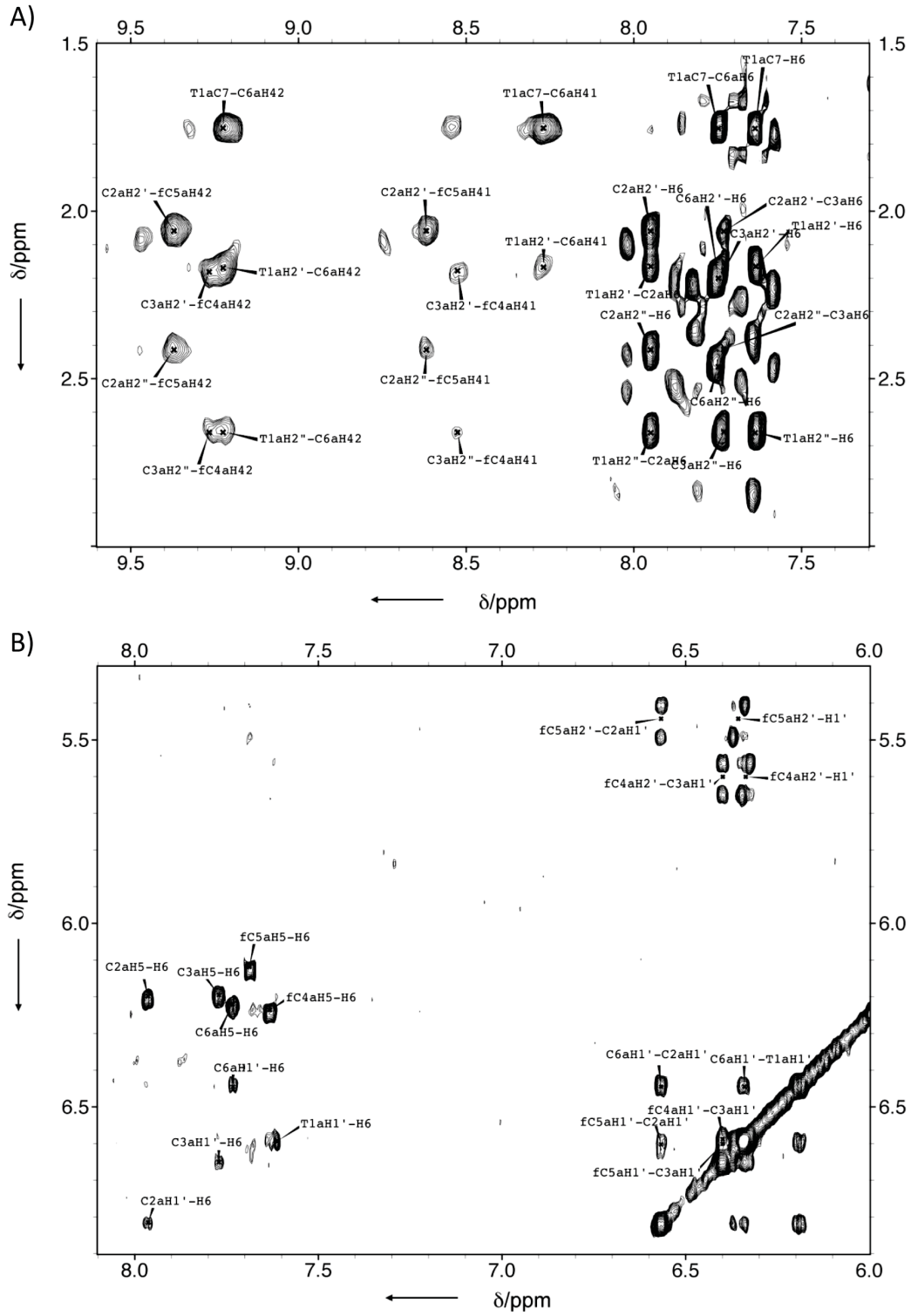


Figure 2.16: NMR spectra of H-4 (5'-TCC(fC)(fC)C-3'). A) Amino-H2'/2" region of the NOESY spectra of H-4 (mixing time = 100 ms, T = 5 °C, pH 6.0). Only peaks corresponding to the major species are labeled. B) H1'–aromatic and H1'-H1' regions of a NOESY spectra in D_2O (mixing time = 250 ms, T = 25 °C, pH 5.0).

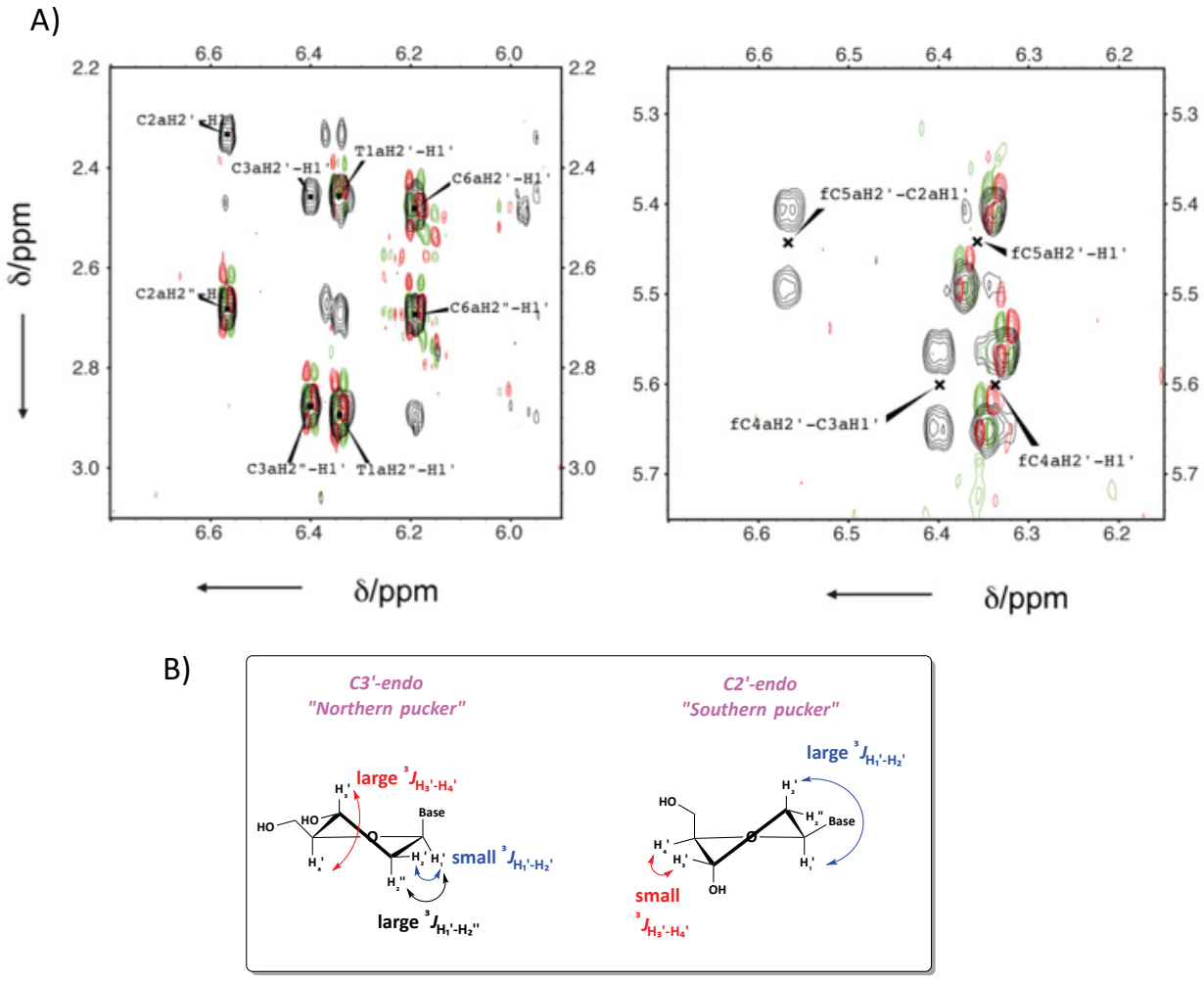


Figure 2.17: A) DQF-COSY (green and red) superimposed on NOESY (black) for H-4 5'-TCC(fC)(fC)C-3', pH 5.0. B) Scheme representing the North and South sugar puckers along with the expected coupling constants.

The structures provide evidence that 2'F-araC fits very well in the otherwise standard 3'E i-motif structure with the stacking order T1-C6-C2-fC5-C3-fC4 (**Figures 2.18** and **AII.5**). The fluorine atoms in the S-type arabinoses point toward the major groove of the i-motif without affecting sugar-sugar contacts critical for i-motif stability (Figure **AII.5**). Geometrical parameters are shown in **Table AII.4**.

The major i-motif species at acidic pH was not the most stable one at neutral pH (**Figure 2.19**). NMR data indicate that the terminal deoxycytidines (C6) were not protonated and were disordered at pH 7.0. Some key NOEs, such as those shown in **Figures 2.20** and **AII.6**, indicated

that the intercalation order is T1-fC5-C2-fC4-C3 (**Figure 2.18**), which is different than that of the major conformer at acidic pH.

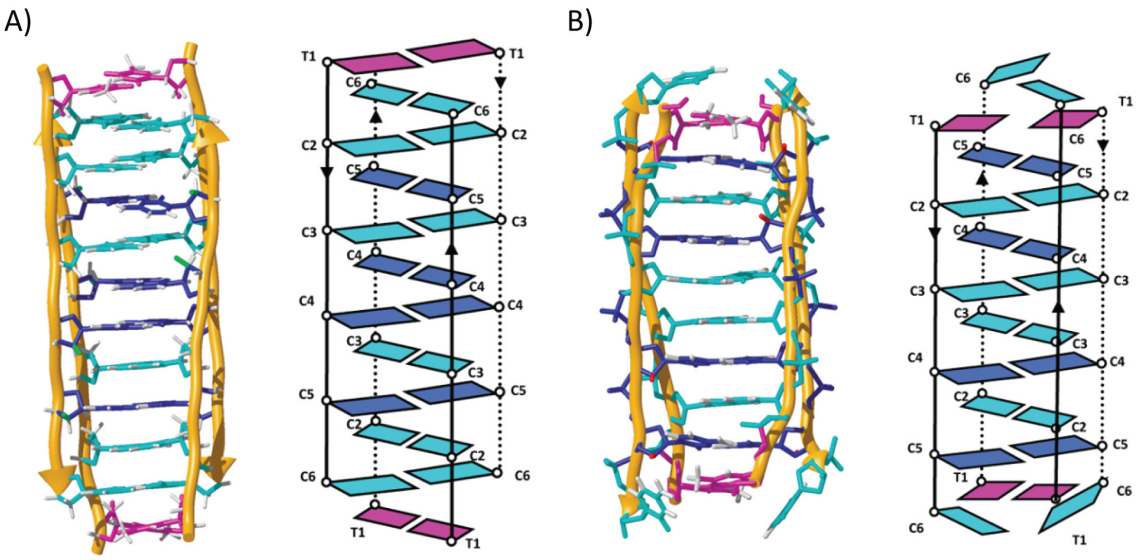


Figure 2.18: A) Average solution structure and schematic representation of the major 3'E species of H-4 at pH 5.0. B) Schematic representation of the major 3'E species of H-4 at pH 7.0. Color code: dC: cyan, 2'F-araC: blue, dT: magenta; and fluorine atoms: green.

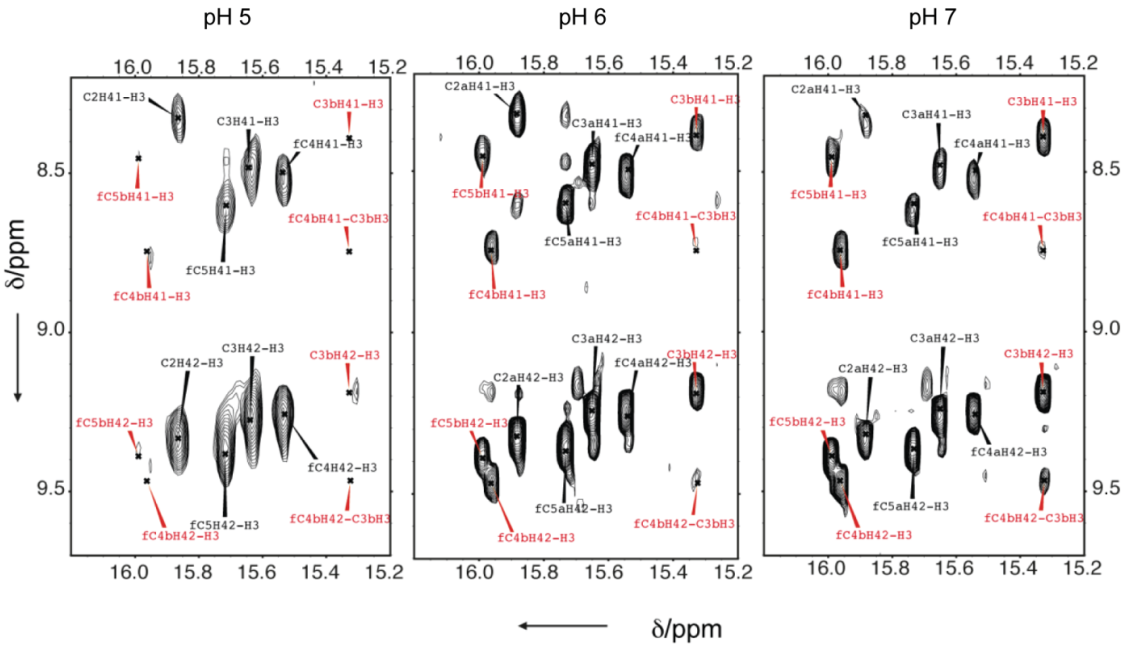


Figure 2.19: Imino-amino region of the NOESY spectra of H-4 (2.53 mM) at different pHs (mixing time = 100 ms, T = 5 °C). Black cross-peaks correspond to the major species at pH 5.0. Red cross-peaks correspond to the major species at pH 7.0.

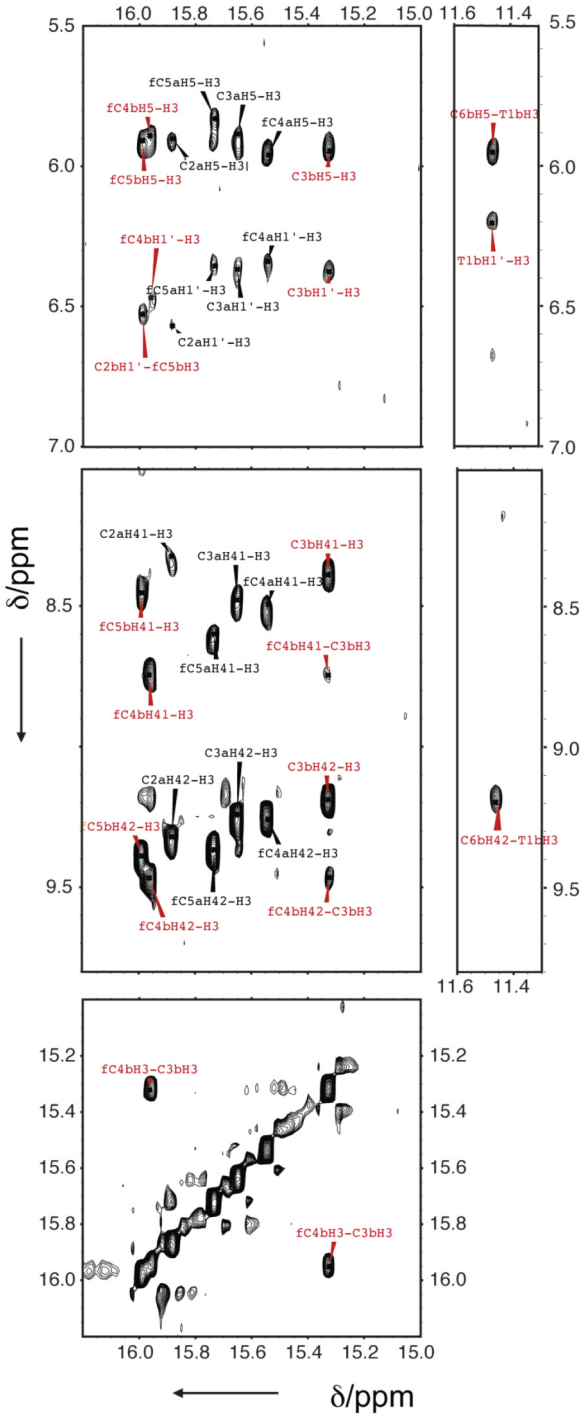


Figure 2.20: NMR spectra H-4 at neural pH. Several regions of the NOESY spectra of H-4 involving exchangeable protons (mixing time = 100 ms, T = 5 °C, pH 7). Cross-peaks labeled in black and letter "a" correspond to the major species at acidic pH. Cross-peaks labeled in red and letter "b" correspond to the major species at neutral pH. NOEs involving H3 of T1 are only observed at neutral pH.

Although the complete sequential assignment of the NMR spectra of the major conformer at neutral pH could not be carried out, a model structure consistent with the NOEs that could be assigned unambiguously was built and refined with molecular dynamics methods (**Figures 2.18** and **AII.6**). Similar to the conformer observed at acidic pH, the neutral pH conformer maintained a 3'E topology with one fewer $C{\cdot}C^+$, due to deprotonation of the terminal $C{\cdot}C^+$ base pair. Thus, disruption of $C6{\cdot}C6^+$ base pair upon raising the pH provokes a rearrangement of the intercalation order, preserving the 3'E topology. This reflects the well-documented higher stability of 3'E *versus* 5'E intermolecular i-motif topologies.[43,44]

2.7 Significance of Results

Our groups and others have extensively studied the stabilizing effect of fluorine substitutions at the sugar C2' position in different nucleic acids motifs.[45-48] 2'F-arabinonucleotides, when incorporated into DNA strands, increase binding affinity towards RNA through formation of non-canonical 2'F···purine (H8) hydrogen bonds.[49,50] Similar interactions have been observed in pure 2'F-ANA·2'F-ANA duplexes.[35] Moreover, the strong electronegativity of fluorine affects the sugar charge distribution (in particular in the H2' proton) and can induce the formation of FC-H···O hydrogen bonds between sequential sugars. This interaction is responsible for the enhanced stability of 2'F-araG substituted G-quadruplexes[30,51] and 2'F-ANA·RNA hybrid duplexes.[29] The structural analysis of H-4 did not provide evidence for FC-H···O hydrogen bonds with the appropriate geometry. However, a number of favorable sequential and inter-strand contacts involving 2'F-araC residues were observed (**Figure 2.21**). Many of these close contacts are facilitated by the south conformation of the 2'F-arabinoses, and hence are absent in unmodified i-motifs in which the sugar conformations are north. In particular, sequential FC-H2'···O4' and inter-strand FC-H2'···O2 distances are much shorter in S-type 2'F-arabinoses. Since fluorine electronegativity provokes a positive charge polarization at the geminal H2' proton, its close contacts with electron-dense oxygen atoms cause favorable electrostatic interactions. Other favorable electrostatic interactions involve sequential F-C-H2'···X (where X = O3', O5') and inter-strand $F{\cdots}H_2N$. In addition to favorable electrostatic interactions, a long-range inductive effect may also strengthen hemiprotonated 2'F-araC·2'F-$araC^+$ base pairs as seen in 2'F-RNA duplexes.[47,52] Stronger base pairing and long-range

electrostatic interactions may also explain the large increase in $pH_{1/2}$ observed for 2'F-araC substituted i-motifs compared to the unmodified structures.

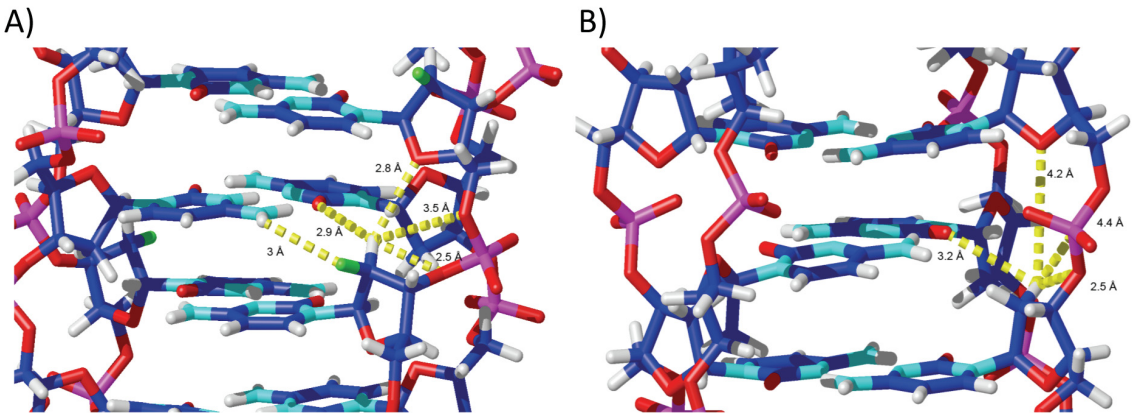


Figure 2.21: Details of sugar contacts in A) 2'F-araC substituted i-motif H-4 and in B) an unmodified i-motif H-1. Close distances corresponding to favorable electrostatic interactions in the substituted i-motif are shown in yellow (the structure is slightly rotated along the Y-axis to better illustrate the relevant distances). The electronegative fluorine and the positively polarized H2' alleviate the unfavorable electrostatic interactions between O2, O3', O4' and O5', characteristic of i-motif structures. Coordinates of the unmodified i-motif are 1YBL.

The possibility of modulating the stability and pH dependence of the i-motif through chemical modification has attracted much attention in recent years. Even though the deoxyribose sugar rings in an i-motif structure adopt a C3'-*endo* conformation, modifications adopting a C3'-*endo* sugar pucker, such as RNA,[27] LNA,[21] and 2'F-RNA,[28] lead to position-dependent stabilization and only at acidic pH values. Our modification is a rather conservative DNA-like modification which preserves the structure of natural i-motifs with only minor alterations. This makes 2'F-araC substituted sequences excellent mimics for the study of molecular recognition processes involving i-motifs. For example, pull-down cell based assays that identify i-motif binding proteins (including potential i-motif specific antibodies) will be significantly facilitated by carrying out these experiments under physiological conditions. The fluorine modification also provides an excellent handle to detect i-motif-ligand interactions *via* ^{19}F-NMR methods. Finally, the nuclease resistance properties conferred by 2'F-araC substitutions are important for *in vivo* applications of i-motif based nanodevices.[14]

2.8 Conclusions

We have shown that i-motifs, including tetrameric sequences and those formed by centromeric and telomeric DNA sequences, can be significantly stabilized by replacing dC units with 2'F-araC residues, by at least 10 kJ/mol per 2'F-araC incorporation. Importantly, replacement of DNA with 2'F-araC does not alter the overall i-motif structure. This provides us with the opportunity to correlate the structural information of transient arrangements of G4 and i-motifs within duplex DNA with transcription factor or polymerase activity under physiological conditions (Chapter 4), and to screen small molecule libraries that bind to our fluorinated i-motifs under physiological conditions.[53]

2.9 Experimental Section

2.9.1 Oligonucleotide Synthesis and Purification

Oligonucleotide synthesis was performed on an ABI 3400 DNA synthesizer (Applied Biosystems) at 1 μmol scale on Unylinker (Chemgenes) CPG solid support. Thymidine (dT), deoxycytidine (*N*-acetyl) (dC), and deoxyadenosine (*N*-Bz) (dA) phosphoramidites were used at 0.1 M concentration in acetonitrile and coupled for 110 s. 2'F-araC was used at 0.13 M concentration and coupled for 600 s. After completion of the synthesis, the CPG was transferred to a 1.5 mL screw-cap eppendorf. Cleavage of the oligonucleotide from the CPG and removal of the nucleobase protecting groups were achieved by treatment with 1000 μL aqueous ammonium hydroxide for 48 hours at room temperature. The ammonia solution was centrifuged and decanted from the CPG. Samples were vented for 30 min, chilled on dry ice and evaporated to dryness. Hexamer sequences were purified by anion exchange HPLC on a Waters 1525 instrument using a Protein-Pak DEAE 5PW column (21.5 mm × 15 cm). The buffer system consisted of water (solution A) and 1 M aqueous lithium perchlorate (solution B), at a flow rate of 4 mL/min. The gradient was 0−40% lithium perchlorate over 50 minutes at 60 °C. Under these conditions the desired peaks eluted at roughly 28 min. The centromeric (HC) and telomeric (HT) sequences were purified by anion exchange HPLC on an Agilent 1200 Series instrument using a Protein-Pak DEAE 5PW column (7.5 x 75 mm) at a flow rate of 1 mL/min using a gradient of 0-24% lithium perchlorate buffer (1.0 M) over 30 minutes at 60 °C. Under these conditions, the desired HC and HT peaks eluted at roughly 21 and 24 min, respectively. Samples were desalted on NAP-25 desalting columns according to manufacturer protocol. The extinction

coefficient of the 2'F-araC was assumed to match that of the unmodified deoxynucleosides. Masses were verified by high resolution LC-MS.

Table 2.5: MS characterization of the studied sequences.

Code	Sequence (5'-3')	*Calculated mass (g/mole)*	*Mass found (g/mole) (-ve mode)*
H-1	dTCCCCC	1688.1375	1687.2140
H-2	dTCC(fC)CC	1706.1280	1705.2034
H-3	dTC(fC)(fC)CC	1724.1185	1723.1887
H-4	dTCC(fC)(fC)C	1724.1185	1723.1913
H-5	dTC(fC)(fC)(fC)C	1742.1090	1741.1887
H-6	dTf(CCCCC)	1778.0990	1777.1629
HC-0	dTCCTTTTCCA	2928.9350	2927.4888
HC-1	dT(fC)CTTTTCCA	2946.9255	2945.4350
HC-2	dT(fC)(fC)TTTTCCA	2964.9160	2963.4230
HC-3	dT(fC)CTTTT(fC)CA	2964.9160	2963.4198
HC-4	dT(fC)(fC)TTTT(fC)(fC)A	3000.8970	2999.4464
HT-0	$d(CCCTAA)_3C_3$	6197.0846	6197.0313
HT-1	$d(C(fC)CTAA)_3C(fC)C$	6269.0466	6269.0313
HT-2	$d(C(fC)(fC)TAA)_3C(fC)(fC)$	6341.0086	6341.0313
HT-3	$d((fC)(fC)CTAA)_3(fC)(fC)C$	6341.0086	6341.0315
HT-4	$d((fC)(fC)(fC)TAA))_3(fC)_3$	6412.9706	6413.0313

2.9.2 UV melting Studies

UV thermal denaturation data was obtained on a Varian CARY 100 UV-visible spectrophotometer equipped with a Peltier temperature controller. The concentration of oligonucleotides used was 4.6 µM for the hexamer sequences and 4 µM for the centromeric and telomeric sequences. Samples were dissolved in appropriate buffer as indicated in the text. Concentrations were determined after quantitating the samples by UV absorbance at $\lambda = 260$ nm. Samples were heated to 90 °C for 15 minutes, then cooled slowly to room temperature, and stored at 5 °C at least 16 h before the measurements were performed. Denaturation curves were acquired at 265 nm at a rate of 0.5 °C/min. Experiments at pH 5.0 and pH 7.0 were performed at least in triplicates at a ramp of 0.5 °C/min and experiments at 0.2 °C/min were performed once for comparison. Samples were kept under a nitrogen sheath at temperatures below 12 °C. The

dissociation temperatures were calculated as the midpoint of the transition ($T_{1/2}$) values using the first derivatives of the experimental data.

2.9.3 Circular Dichroism

Circular Dichroism (CD) studies were performed at 5 °C on a JASCO J-810 spectropolarimeter using a 1 mm path length cuvette. Temperature was maintained using the Peltier unit within the instrument. Spectra were recorded from 350-230 nm at a scan rate of 100 nm/min and a response time of 2.0 s with three acquisitions recorded for each spectrum. The spectra were normalized by subtraction of the background scan with buffer. Data were smoothed using the means-movement function within the JASCO graphing software. Oligonucleotide solutions for CD measurements were prepared with 10 mM sodium phosphate buffer (pH 5.0 and pH 7.0) in a similar manner to that used for UV melting. The concentration of oligonucleotides used was 30 μM for the hexamer sequences, 100 μM for the centromeric sequences, and 50 μM for the telomeric sequences.

2.9.4 Gel Electrophoresis

Native polyacrylamide gel electrophoresis was performed utilizing 24% polyacrylamide at pH 5.0 using TAE (Tris:Acetate:EDTA) buffer. Running buffer was 1x TAE. Samples were equilibrated in 10 mM sodium phosphate (pH 5.0) as described in the UV melting section. The gels were run at 200 V inside a refrigerator (5 °C) for 12 hours or at 280 V for 6 hours. Gels were visualized by treatment with Stains-All dye (Sigma E-9379) or UV shadowing. The oligonucleotide controls were dT_{12} and dT_{24} strands.

2.9.5 NMR Experiments and Constraints

NMR Experiments: Samples for NMR experiments were suspended in 300 μL of either D_2O or 9:1 H_2O/D_2O in 10 mM sodium phosphate buffer. NMR spectra were acquired on Bruker Avance spectrometers operating at 600, 700, or 800 MHz, and processed with Topspin software. TOCSY spectra were recorded with standard MLEV17 spinlock sequence and with 80 ms mixing time. NOESY spectra in H_2O were acquired with 50 and 150 ms mixing times. For 2D experiments in H_2O, water suppression was achieved by including a WATERGATE module in the pulse sequence prior to acquisition.[54] Two-dimensional experiments were carried out at temperatures ranging from 5 to 25 °C. ^{19}F resonances were assigned from ^{1}H-^{19}F HETCOR and

^{19}F detected HOESY spectra.[55] The spectral analysis program Sparky was used for semiautomatic assignment of the NOESY cross-peaks.[56]

Experimental NMR constraints: Distance constraints were obtained from a qualitative estimation of NOE intensities. In addition to these experimentally derived constraints, hydrogen bond and planarity constraints for the base pairs were used in the initial DYANA calculations. Target values for distances and angles related to hydrogen bonds were set to values obtained from crystallographic data in related structures.[57] Due to the relatively broad line-widths of the sugar proton signals, J-coupling constants were not accurately measured, but only roughly estimated from DQF-COSY cross-peaks. Loose values were set for the sugar dihedral angles δ, ν_1 and ν_2 to constrain deoxyribose conformation to North or South domain. No backbone angle constraints were employed. Distance constraints with their corresponding error bounds were incorporated into the AMBER potential energy by defining a flat-well potential term.

2.9.6 Structural Determination

Structures were calculated with the program DYANA 1.4[58] and further refined with the SANDER module of the molecular dynamics package AMBER 7.0.[59] The resulting DYANA structures were used as starting points for the AMBER refinement, consisting of 1 ns trajectories in which explicit solvent molecules were included and using the Particle Mesh Ewald method to evaluate long-range electrostatic interactions. Non-experimental constraints used in the initial DYANA calculations were removed in the AMBER refinement. The specific protocols for these calculations have been described elsewhere.[42] The AMBER-98 force-field was used to describe the DNA, and the TIP3P model was used to simulate water molecules.[60] Analysis of the representative structures as well as the MD trajectories was carried out with the programs Curves V5.1[61] and MOLMOL.[62]

2.9.7 Differential Scanning Calorimetry

DSC measurements were performed using a NanoDSC-III (TA Instruments, USA). All samples were prepared as 150 μM solutions in 10 mM sodium phosphate buffer (pH 5.0). Sample data were collected in triplicate by scanning from 5-80 °C at a scan rate of 0.5 °C per minute. Fitting procedures and statistical analysis methods are described in detail in **Appendix AII.1**. Errors were calculated according to the variance-covariance method.[33]

2.10 References

1. Choi, J.; Majima, T., Conformational changes of non-B DNA. *Chem. Soc. Rev.* **2011,** *40* (12), 5893-5909.
2. Sen, D.; Gilbert, W., Formation of Parallel 4-Stranded Complexes by Guanine-Rich Motifs in DNA and Its Implications for Meiosis. *Nature* **1988,** *334* (6180), 364-366.
3. Moye, A. L.; Porter, K. C.; Cohen, S. B.; Phan, T.; Zyner, K. G.; Sasaki, N.; Lovrecz, G. O.; Beck, J. L.; Bryan, T. M., Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nat. Commun.* **2015**, *6*, 1-12
4. Gehring, K.; Leroy, J. L.; Gueron, M., A Tetrameric DNA-Structure with Protonated Cytosine.Cytosine Base-Pairs. *Nature* **1993,** *363* (6429), 561-565.
5. Benabou, S.; Aviñó, A.; Eritja, R.; González, C.; Gargallo, R., Fundamental aspects of the nucleic acid i-motif structures. *Rsc Adv.* **2014,** *4* (51), 26956.
6. Day, H. A.; Pavlou, P.; Waller, Z. A., i-Motif DNA: structure, stability and targeting with ligands. *Bioorg. Med. Chem.* **2014,** *22* (16), 4407-18.
7. Garavis, M.; Escaja, N.; Gabelica, V.; Villasante, A.; Gonzalez, C., Centromeric Alpha-Satellite DNA Adopts Dimeric i-Motif Structures Capped by AT Hoogsteen Base Pairs. *Chem. Eur. J.* **2015,** *21* (27), 9816-24.
8. Garavis, M.; Mendez-Lago, M.; Gabelica, V.; Whitehead, S. L.; Gonzalez, 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*, 1-10.
9. Phan, A. T.; Gueron, 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.
10. Kang, C. H.; Berger, I.; Lockshin, C.; Ratliff, R.; Moyzis, R.; Rich, A., Crystal-Structure of Intercalated 4-Stranded D(C3t) at 1.4 Angstrom Resolution. *Proc. Natl. Acad. Sci. USA* **1994,** *91* (24), 11636-11640.
11. 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. *J. Am. Chem. Soc.* **2014,** *136* (11), 4172-4185.
12. Kendrick, S.; Kang, H. J.; Alam, M. P.; Madathil, M. M.; Agrawal, P.; Gokhale, V.; Yang, D. Z.; 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.
13. Guittet, E.; Renciuk, D.; Leroy, J. L., Junctions between i-motif tetramers in supramolecular structures. *Nucleic Acids Res.* **2012,** *40* (11), 5162-70.
14. Dong, Y. C.; Yang, Z. Q.; Liu, D. S., DNA Nanotechnology Based on i-Motif Structures. *Acc. Chem. Res.* **2014,** *47* (6), 1853-1860.

15. Li, T.; Famulok, M., I-motif-programmed functionalization of DNA nanocircles. *J. Am. Chem. Soc.* **2013,** *135* (4), 1593-9.
16. Berger, I.; Egli, M.; Rich, A., Inter-strand C-H···O hydrogen bonds stabilizing four-stranded intercalated molecules: Stereoelectronic effects of 04' in cytosine-rich DNA. *Proc. Natl. Acad. Sci. USA* **1996,** *93* (22), 12116-12121.
17. 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.
18. Kendrick, S.; Hurley, L. H., The role of G-quadruplex/i-motif secondary structures as cis-acting regulatory elements. *Pure Appl. Chem.* **2010,** *82* (8), 1609-1621.
19. 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), 12751-67.
20. Day, H. A.; Huguin, C.; Waller, Z. A. E., Silver cations fold i-motif at neutral pH. *Chem. Commun.* **2013,** *49* (70), 7696-7698.
21. Kumar, N.; Nielsen, J. T.; Maiti, S.; Petersen, M., i-Motif formation with locked nucleic acid (LNA). *Angew. Chem.* **2007,** *46* (48), 9220-2.
22. 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), 752-5.
23. Perez-Rentero, S.; Gargallo, R.; Gonzalez, C.; Eritja, R., Modulation of the stability of i-motif structures using an acyclic threoninol cytidine derivative. *Rsc Adv.* **2015,** *5* (78), 63278-63281.
24. Kanaori, K.; Sakamoto, S.; Yoshida, H.; Guga, P.; Stec, W.; Tajima, K.; Makino, K., Effect of phosphorothioate chirality on i-motif structure and stability. *Biochemistry* **2004,** *43* (19), 5672-5679.
25. Krishnan-Ghosh, Y.; Stephens, E.; Balasubramanian, S., PNA forms an i-motif. *Chem. Commun.* **2005,** (42), 5278-80.
26. 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), 529-35.
27. 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.
28. Fenna, C. P.; Wilkinson, V. J.; Arnold, J. R.; Cosstick, R.; Fisher, J., The effect of 2'-fluorine substitutions on DNA i-motif conformation and stability. *Chem. Commun.* **2008,** (30), 3567-9.
29. Martin-Pintado, N.; Deleavey, G. F.; Portella, G.; Campos-Olivas, R.; Orozco, M.; Damha, M. J.; Gonzalez, C., Backbone FC-H...O hydrogen bonds in 2'F-substituted nucleic acids. *Angew. Chem.* **2013,** *52* (46), 12065-8.

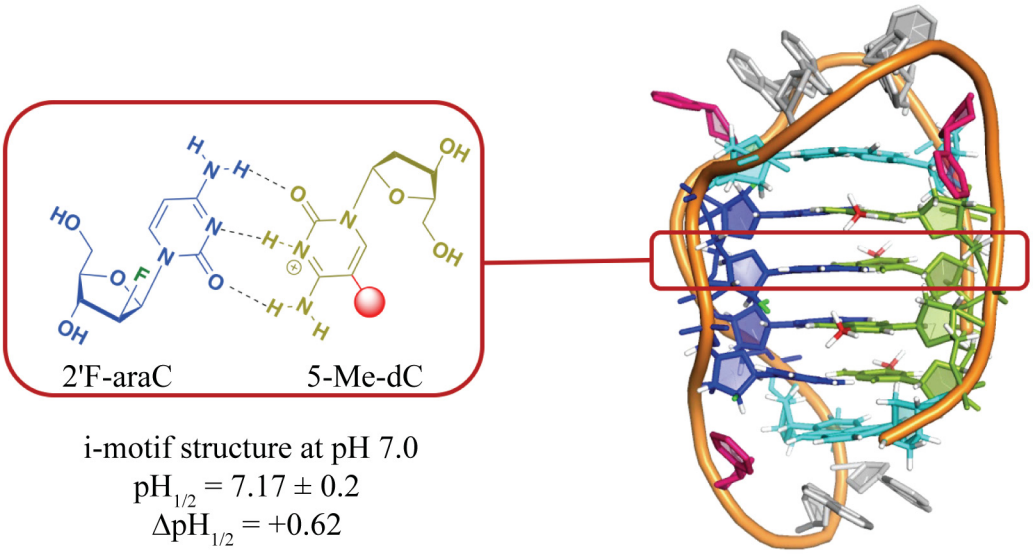
30. Martin-Pintado, N.; Yahyaee-Anzahaee, M.; Deleavey, G. F.; Portella, G.; Orozco, M.; Damha, M. J.; Gonzalez, C., Dramatic effect of furanose C2' substitution on structure and stability: directing the folding of the human telomeric quadruplex with a single fluorine atom. *J. Am. Chem. Soc.* **2013,** *135* (14), 5344-7.
31. Kanehara, H.; Mizuguchi, M.; Tajima, K.; Kanaori, K.; Makino, K., Spectroscopic evidence for the formation of four-stranded solution structure of oligodeoxycytidine phosphorothioate. *Biochemistry* **1997,** *36* (7), 1790-1797.
32. 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.
33. Tellinghuisen, J., Statistical error propagation. *J. Phys. Chem. A* **2001,** *105* (15), 3917-3921.
34. Kreutz, C.; Kahlig, H.; Konrat, R.; Micura, R., Ribose 2 '-F labeling: A simple tool for the characterization of RNA secondary structure equilibria by F-19 NMR spectroscopy. *J. Am. Chem. Soc.* **2005,** *127* (33), 11558-11559.
35. Martin-Pintado, N.; Yahyaee-Anzahaee, M.; Campos-Olivas, R.; Noronha, A. M.; Wilds, C. J.; Damha, M. J.; Gonzalez, C., The solution structure of double helical arabino nucleic acids (ANA and 2 ' F-ANA): effect of arabinoses in duplex-hairpin interconversion. *Nucleic Acids Res.* **2012,** *40* (18), 9329-9339.
36. Luo, J.; Sarma, M. H.; Yuan, R. D.; Sarma, R. H., Nmr-Study of Self-Paired Parallel Duplex of D(Aaaaaccccc) in Solution. *Febs Lett.* **1992,** *306* (2-3), 223-228.
37. Geinguenaud, F.; Liquier, J.; Brevnov, M. G.; Petrauskene, O. V.; Alexeev, Y. I.; Gromova, E. S.; Taillandier, E., Parallel self-associated structures formed by T,C-rich sequences at acidic pH. *Biochemistry* **2000,** *39* (41), 12650-12658.
38. Lim, K. W.; Lacroix, L.; Yue, D. J. E.; Lim, J. K. C.; Lim, J. M. W.; Phan, A. T., Coexistence of Two Distinct G-Quadruplex Conformations in the hTERT Promoter. *J. Am. Chem. Soc.* **2010,** *132* (35), 12331-12342.
39. Patel, D. J.; Kozlowski, S. A.; Marky, L. A.; Broka, C.; Rice, J. A.; Itakura, K.; Breslauer, K. J., Premelting and Melting Transitions in the D(Cgcgaattcgcg) Self-Complementary Duplex in Solution. *Biochemistry* **1982,** *21* (3), 428-436.
40. Rule, G. S.; Hitchens, T. K., *Fundamentals of Protein NMR Spectroscopy*. Springer Netherlands: 2006.
41. Wemmer, D. E.; Chou, S. H.; Hare, D. R.; Reid, B. R., Duplex-Hairpin Transitions in DNA - Nmr-Studies on Cgcgtatacgcg. *Nucleic Acids Res.* **1985,** *13* (10), 3755-3772.
42. Soliva, R.; Monaco, V.; Gomez-Pinto, I.; Meeuwenoord, N. J.; Van der Marel, G. A.; Van Boom, J. H.; Gonzalez, C.; Orozco, M., Solution structure of a DNA duplex with a chiral alkyl phosphonate moiety. *Nucleic Acids Res.* **2001,** *29* (14), 2973-2985.
43. Leroy, J. L.; Snoussi, K.; Gueron, M., Investigation of the energetics of C-H···O hydrogen bonds in the DNA i-motif via the equilibrium between alternative intercalation topologies. *Magn. Reson. Chem.* **2001,** *39*, S171-S176.

44. 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.
45. Lech, C. J.; Li, Z.; Heddi, B.; Phan, A. T., 2'-F-ANA-guanosine and 2'-F-guanosine as powerful tools for structural manipulation of G-quadruplexes. *Chem. Commun.* **2012,** *48* (93), 11425-7.
46. Li, Z.; Lech, C. J.; Phan, A. T., Sugar-modified G-quadruplexes: effects of LNA-, 2'F-RNA- and 2'F-ANA-guanosine chemistries on G-quadruplex structure and stability. *Nucleic Acids Res.* **2014,** *42* (6), 4068-79.
47. Pallan, P. S.; Greene, E. M.; Jicman, P. A.; Pandey, R. K.; Manoharan, M.; Rozners, E.; Egli, M., Unexpected origins of the enhanced pairing affinity of 2 '-fluoro-modified RNA. *Nucleic Acids Res.* **2011,** *39* (8), 3482-3495.
48. Manoharan, M.; Akinc, A.; Pandey, R. K.; Qin, J.; Hadwiger, P.; John, M.; Mills, K.; Charisse, K.; Maier, M. A.; Nechev, L.; Greene, E. M.; Pallan, P. S.; Rozners, E.; Rajeev, K. G.; Egli, M., Unique Gene-Silencing and Structural Properties of 2 '-Fluoro-Modified siRNAs. *Angew. Chem. Int. Edit.* **2011,** *50* (10), 2284-2288.
49. Watts, J. K.; Martin-Pintado, N.; Gomez-Pinto, I.; Schwartzentruber, J.; Portella, G.; Orozco, M.; Gonzalez, C.; Damha, M. J., Differential stability of 2 ' F-ANA.RNA and ANA.RNA hybrid duplexes: roles of structure, pseudohydrogen bonding, hydration, ion uptake and flexibility. *Nucleic Acids Res.* **2010,** *38* (7), 2498-2511.
50. Anzahaee, M. Y.; Watts, J. K.; Alla, N. R.; Nicholson, A. W.; Damha, M. J., Energetically Important C-H···F-C Pseudohydrogen Bonding in Water: Evidence and Application to Rational Design of Oligonucleotides with High Binding Affinity. *J. Am. Chem. Soc.* **2011,** *133* (4), 728-731.
51. Peng, C. G.; Damha, M. J., G-quadruplex induced stabilization by 2'-deoxy-2'-fluoro-D-arabinonucleic acids (2'F-ANA). *Nucleic Acids Res.* **2007,** *35* (15), 4977-88.
52. Patra, A.; Paolillo, M.; Charisse, K.; Manoharan, M.; Rozners, E.; Egli, M., 2 '-Fluoro RNA Shows Increased Watson-Crick H-Bonding Strength and Stacking Relative to RNA: Evidence from NMR and Thermodynamic Data. *Angew. Chem. Int. Edit.* **2012,** *51* (47), 11863-11866.
53. Stephenson, A. W. I.; Partridge, A. C.; Filichev, V. V., Synthesis of beta-Pyrrolic-Modified Porphyrins and Their Incorporation into DNA. *Chem. Eur. J.* **2011,** *17* (22), 6227-6238.
54. Piotto, M.; Saudek, V.; Sklenar, V., Gradient-Tailored Excitation for Single-Quantum Nmr-Spectroscopy of Aqueous-Solutions. *J. Biomol. NMR* **1992,** *2* (6), 661-665.
55. Yu, C.; Levy, G. C., Two-Dimensional Heteronuclear Noe (Hoesy) Experiments - Investigation of Dipolar Interactions between Heteronuclei and Nearby Protons. *J. Am. Chem. Soc.* **1984,** *106* (22), 6533-6537.
56. Goddard, D. T.; Kneller, D. G. *SPARKY*, 3rd edn, University of California, San Francisco.

57. Cai, L.; Chen, L. Q.; Raghavan, S.; Ratliff, R.; Moyzis, R.; Rich, A., Intercalated cytosine motif and novel adenine clusters in the crystal structure of the Tetrahymena telomere. *Nucleic Acids Res.* **1998,** *26* (20), 4696-4705.
58. Guntert, P.; Mumenthaler, C.; Wuthrich, K., Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **1997,** *273* (1), 283-298.
59. Case, D. A.; Pearlman, D. A.; Caldwell, J. W. I. III AMBER7, University of California, San Francisco, **2007**.
60. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *J. Am. Chem. Soc.* **1995,** *117* (19), 5179-5197.
61. Lavery, R.; Sklenar, H. *CURVES, helical analysis of irregular nucleic acids. 3.0.*, Laboratory of Theoretical Biochemistry CNRS, Paris, **1990**.
62. Koradi, R.; Billeter, M.; Wuthrich, K., MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graphics* **1996,** *14* (1), 29-55.

Chapter 3

Probing Synergistic Effects of DNA Methylation and 2'-β-Fluorination on i-Motif Stability



i-motif structure at pH 7.0
$pH_{1/2}$ = 7.17 ± 0.2
$\Delta pH_{1/2}$ = +0.62

The majority of this chapter is reproduced from: "Probing synergistic effects of DNA methylation and 2'-β-fluorination on i-motif stability", **Hala Abou Assi**, Yu Chen Lin, Israel Serrano, Carlos González, and Masad J. Damha, *Chemistry – A European Journal*, **2017**, doi: 10.1002/chem.201704591.

"Good is the enemy of great…"

–James C. Collins

3.1 Introduction

As discussed in the previous chapters of this thesis, DNA sequences containing stretches of cytosine residues can form intercalated, quadruple-helical i-motif structures under acidic conditions, and consist of two parallel-stranded duplexes that associate head-to-tail through $C{\cdot}C^+$ base pair intercalation.[1]

Numerous strategies have been developed to promote i-motif formation at physiological pH, in order to assist in identifying ligands which can selectively interact with i-motifs *in vivo*[2] and determine their role in the regulation of transcription[3] and in telomere biology.[4-10] Recently pursued strategies include extending the length of the C-tracts;[5,6] altering the sugar, nucleobase, or phosphate moiety;[11] utilizing branched RNA to hold the intercalated duplexes together;[12] introducing the i-motif-forming sequence into a supercoiled DNA plasmid;[7] and utilizing single-walled carbon nanotubes,[13] graphene quantum dots,[14] or small molecule ligands.[15] Furthermore, a wide range of modified sugars has been investigated in i-motif formation, including RNA,[16-18] arabinonucleic acid (ANA),[16] locked (LNA)[19] and unlocked (UNA) nucleic acids,[20,21] 2'F-RNA,[22] and 2'-deoxy-2'-fluoro-arabinonucleic acids (2'F-ANA).[23] Of these, only our 2'F-araC modification has been found to significantly stabilize i-motif structures at neutral pH (**Figure 3.1**).[23] The nucleobase modifications studied include 5-fluoro, 5-bromo, 5-iodo, and 5-methylcytosine.[24] The latter stabilizes i-motif structures up to pH 6.0 ($\Delta T_{1/2} \approx +1.9$ ^{o}C; $\Delta pH_{1/2} \approx +0.11$, relative to cytosine) (**Figure 3.1**).[25]

5-Methylcytidine (5-Me-dC) is of particular interest as C-rich regions in genomic DNA located near transcriptional start sites (TSS)[26] are susceptible to methylation and 5-hydroxymethylation (5-hMe-dC).[27,28] In fact, 5-Me-dC and 5-hMe-dC are considered epigenetic regulators of gene expression.[29-31] Furthermore, 5-Me-dC plays a significant role in genomic imprinting, transcriptional inhibition, and gene silencing applications.[32,33] The presence of 5-Me-dC in DNA telomeric sequences affects telomere length, telomerase activity, and telomeric repeat-containing RNA (TERRA) transcription.[34,35] Therefore, investigating the effect of 5-Me-dC modification on the stability of telomeric i-motifs is of great interest.[24,25]

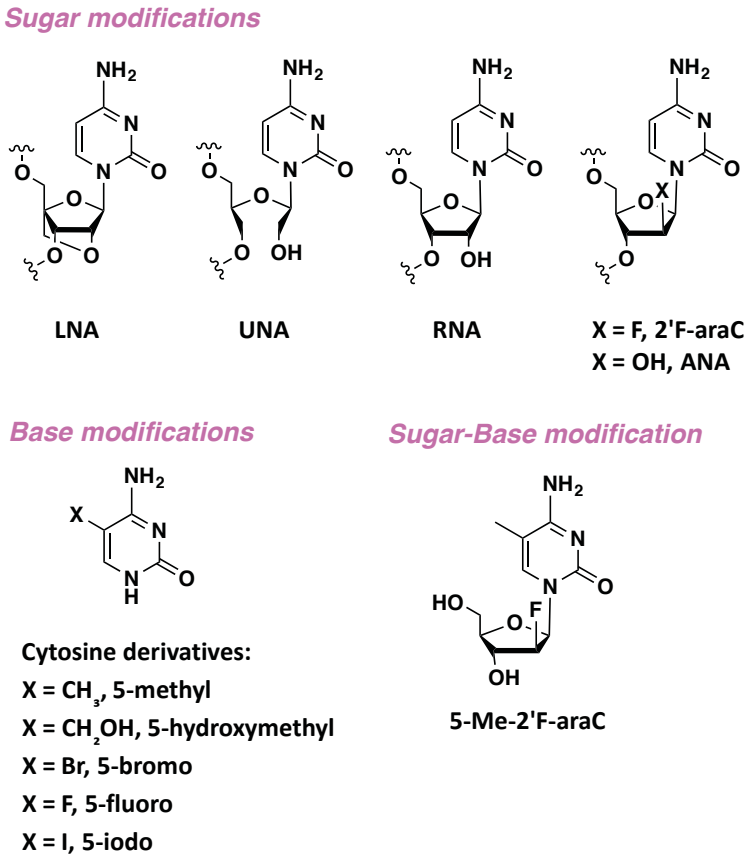


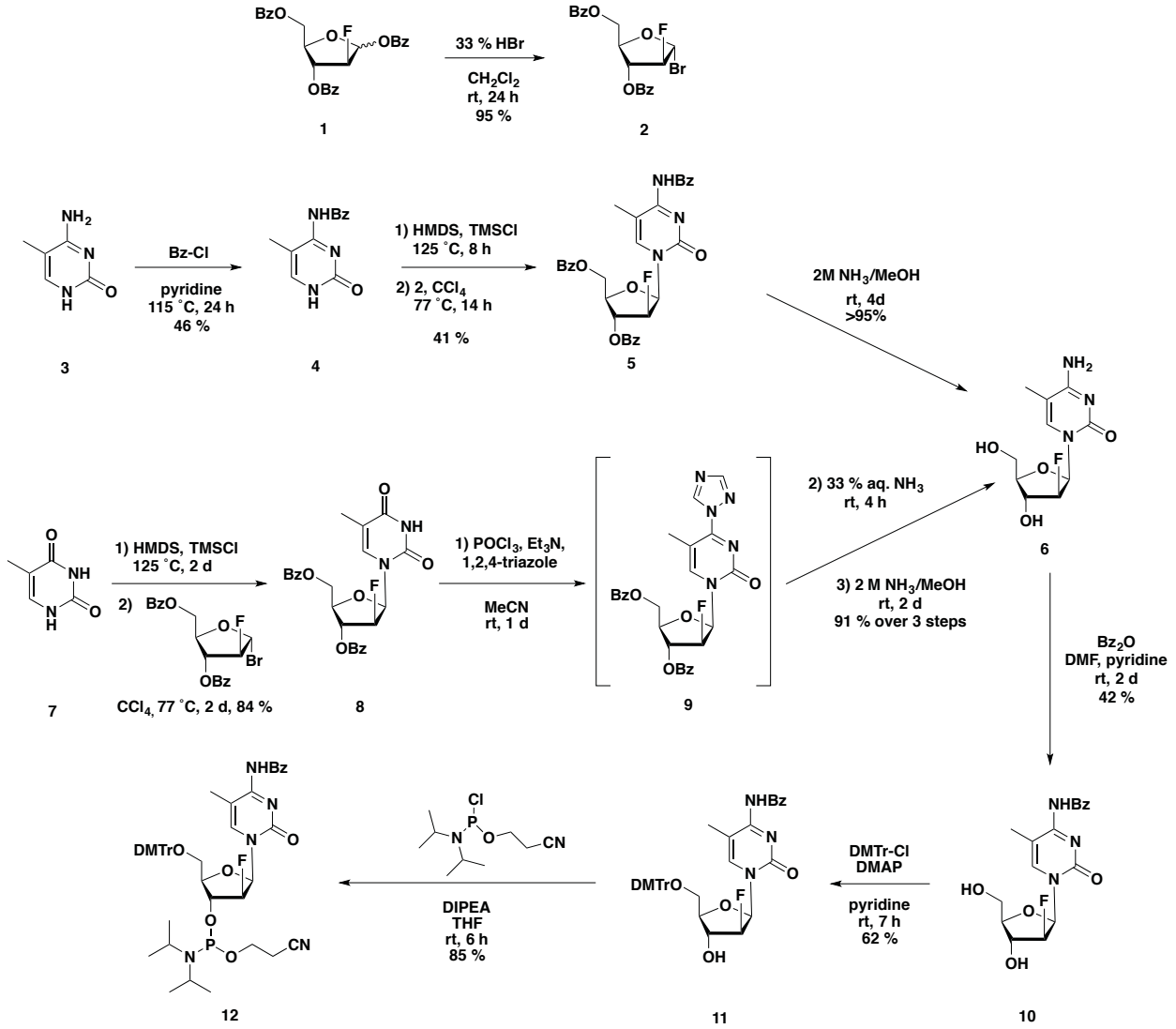
Figure 3.1: Base and sugar modifications previously introduced in i-motif structures.

Given the known stabilizing effect of 5-methylcytosine and 2'F-arabinose in i-motifs, the aim of this chapter was to synthesize 2'-deoxy-2'-fluoro-5-methyl-arabinocytidine (5-Me-2'F-araC) and study its impact on i-motif structure and stability. A second aim was to assess potential synergistic effects of several cytidine derivatives (5-Me-2'F-araC (**fC**^{m}), 2'F-araC (**fC**), 5-Me-dC (**C**^{m}), and dC (C)) and identify combinations of these that provide optimal stability at physiological pH.

3.2 Synthesis of 2'-Deoxy-2'-Fluoro-5-Methyl-Arabinocytidine Phosphoramidite

The synthesis of the novel phosphoramidite **12** is outlined in **Scheme 3.1**. The brominated sugar precursor **2** was prepared from the readily available benzoylated sugar **1** following previously published protocols.[36] The first synthetic route we pursued required benzoyl protection of 5-methylcytosine (**3**) to afford **4** in 46% yield. Glycosylation was accomplished by silylation of the nucleobase **4**, followed by coupling with the brominated sugar

in carbon tetrachloride (CCl_4), resulting in an α/β-anomeric mixture. The anomeric mixture was separated by column chromatography to afford the fully protected β-nucleoside **5** in 41% yield.



Scheme 3.1: Synthetic scheme representing the synthesis of 5-Me-2'F-araC phosphoramidite **12**.

To overcome the challenge of separating α/β-anomeric mixtures, an alternative synthetic route was devised. Here, glycosylation of thymine (**7**) with the brominated sugar precursor **2** afforded the protected β-thymidine nucleoside (**8**) with anomeric purity and high yields (**Scheme 3.1**). Thymidine (**8**) was first converted to 5-Me-cytidine (**6**) using the triazole intermediate method. Reaction of **8** with phosphoryl chloride, triethylamine, and triazole provided intermediate **9**, which was monitored by fluorescence on thin layer chromatography and

followed by removal of the benzoyl groups to give nucleoside **6** in 91% yield over three steps. The NOESY NMR spectra of **6** showed a clear correlation between H-1' and H-4' of the sugar ring and between H-5' of the sugar and H-6 of the nucleobase, as expected for the β-anomer (**Figure 3.2**). Selective benzoylation of the N^4 position gave **10**, which upon 5'-dimethoxytritylation and 3'-phosphitylation afforded the desired 5-Me-2'F-araC phosphoramidite derivative **12** as a mixture of diastereomers (^{31}P NMR, **Figure 3.3**). The long-range ^{19}F-^{31}P coupling ($J_{F\text{-}P}$ = 7.8-9.4 Hz) was consistent with the "W"-like conformation adopted by the F2'-C2'-C3'-O3'-P3' framework in a C2'-*endo* arabinose pucker.[37]

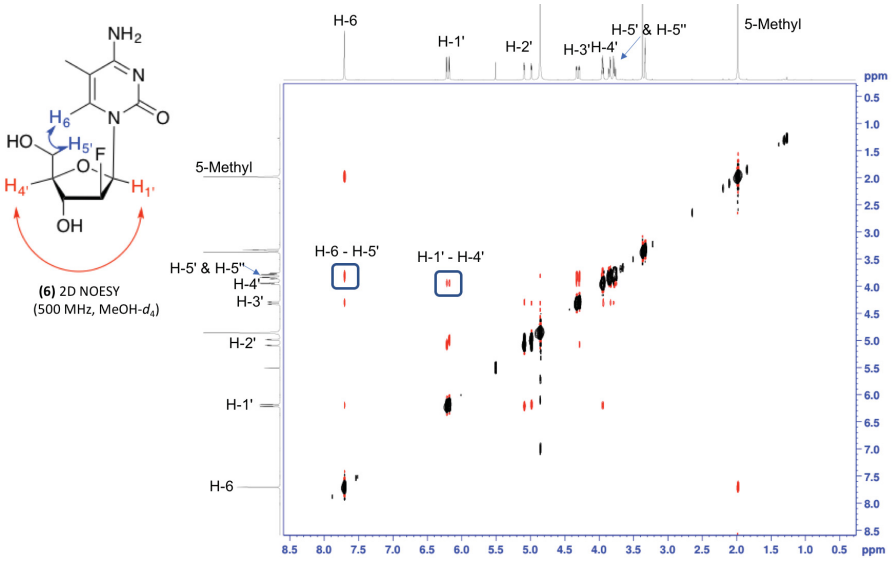


Figure 3.2: NOESY NMR spectrum of 5-methyl-1-[2-deoxy-2-fluoro-β-D-arabinofuranosyl]cytosine (**6**).

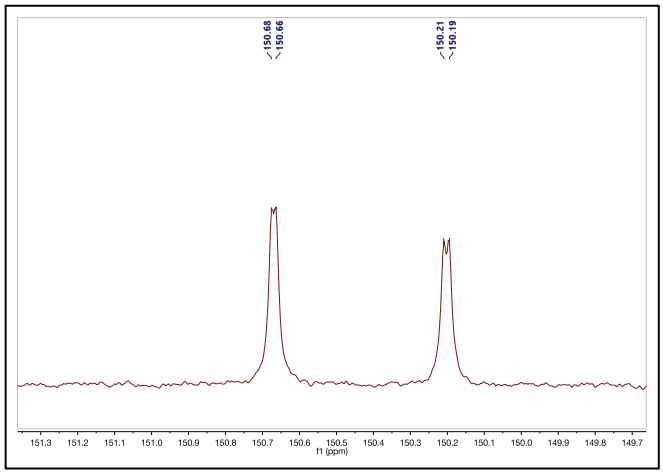


Figure 3.3: ^{31}P-NMR spectrum of amidite **12** showing the long-range ^{19}F-^{31}P coupling.

3.3 Studying the Effect of Nucleobase and Sugar Modifications on i-Motif Formation

A library of eleven oligonucleotides (ONs) was synthesized; their sequences and properties are listed in **Tables 3.1**. i-Motif formation and stability were investigated through circular dichroism (CD), melting temperature ($T_{1/2}$) experiments, and in the case of HJ-2, structural determination *via* NMR. The $T_{1/2}$ and $pH_{1/2}$ values of the studied ONs were very close to the physiological values. In fact, all the sequences studied were more stable than the control strand (HT-0), at pH 5.0, 7.0, and 7.4 (**Tables 3.1** and **Figure 3.4**).

Table 3.1: UV-melting data (°C) of the modified oligonucleotide sequences.

Code	Sequence (5' - 3')	$T_{1/2}$ 5.0	$\Delta T_{1/2}$ 5.0	$T_{1/2}$ 7.0	$\Delta T_{1/2}$ 7.0	$T_{1/2}$ 7.4	Base pairing
HT-0	CCC TAA CCC TAA CCC TAA CCC	59.2 ±0.2	-	15.2 ±0.6	-	-	$C \cdot C^+$
HT-2	**fC**fC TAA C**fC**fC TAA C**fC**fC TAA C**fC**fC	74.1 ±0.5	+14.9	32.1 ±0.3	+16.9	27.3 ±0.2	**fC**$\cdot$**fC**$^+$
HTM-1	C**C^m**C TAA C**C^m**C TAA C**C^m**C TAA C**C^m**C	64.5	+5.3	19.9	+4.7	n.d.	**C^m**$\cdot$**C^{m+}**
HTM-2	C**C^{m}C^m** TAA C**C^{m}C^m** TAA C**C^{m}C^m** TAA C**C^{m}C^m**	64.8 ±0.2	+5.6	31.0 ±0.1	+15.8	27.6 ±0.1	**C^m**$\cdot$**C^{m+}**
HTM-3	**C^{m}C^{m}C** TAA **C^{m}C^{m}C** TAA **C^{m}C^{m}C** TAA **C^{m}C^{m}C**	63.6	+4.4	17.7	+2.5	n.d.	**C^m**$\cdot$**C^{m+}**
HTM-4	**C^{m}C^{m}C^m** TAA **C^{m}C^{m}C^m** TAA **C^{m}C^{m}C^m** TAA **C^{m}C^{m}C^m**	66.2	+7.0	22.7	+7.5	n.d.	**C^m**$\cdot$**C^{m+}**
HTFM	C**fC^{m}fC^m** TAA C**fC^{m}fC^m** TAA C**fC^{m}fC^m** TAA C**fC^{m}fC^m**	70.4 ±0.5	+11.2	30.3 ±0.2	+15.1	25.2 ±0.2	**fC^m**$\cdot$**fC^{m+}**
HJ-1	C**fC**C^m TAA C**C^{m}fC** TAA C**fC**C^m TAA C**C^{m}fC**	69.9 ±0.2	+10.7	31.2 ±0.2	+16.0	27.9 ±0.1	**fC**$\cdot$**fC**$^+$ **C^m**$\cdot$**C^{m+}**
HJ-2	C**C^{m}C^m** TAA C**fC**fC TAA C**fC**fC TAA C**C^{m}C^m**	75.5 ±0.2	+16.3	35.0 ±0.2	+19.8	32.2 ±0.0	**fC**$\cdot$**C^{m+}**
HJ-3	C**fC^{m}fC^m** TAA C**fC**fC TAA C**fC**fC TAA C**fC^{m}fC^m**	74.5 ±0.3	+15.3	33.0 ±0.1	+17.8	29.5 ±0.1	**fC^m**$\cdot$**fC**$^+$
HJ-4	C**fC^{m}fC^m** TAA CCC TAA CCC TAA C**fC^{m}fC^m**	68.8 ±0.2	+9.6	28.7 ±0.6	+13.5	25.3 ±0.3	**fC^m**$\cdot$C$^+$

Oligonucleotide concentration: 4 µM (single strands). $T_{1/2}$ data were calculated from UV-visible spectroscopy thermal denaturation profiles where $T_{1/2}$ corresponds to the midpoint of the dissociation transition obtained at 0.5 °C /min. $\Delta T_{1/2}$ values were calculated for the main melting transitions relative to the respective unmodified strands. (-) indicates sequences for which no melting transition was detected, or an absence of $\Delta T_{1/2}$; (n.d.) indicates not determined. Nucleotide codes: **fC**: 2'F-araC, **C^m**: 5-Me-dC, **fC^m**: 5-Me-2'F-araC; and C, T, and A: dC/A/T).

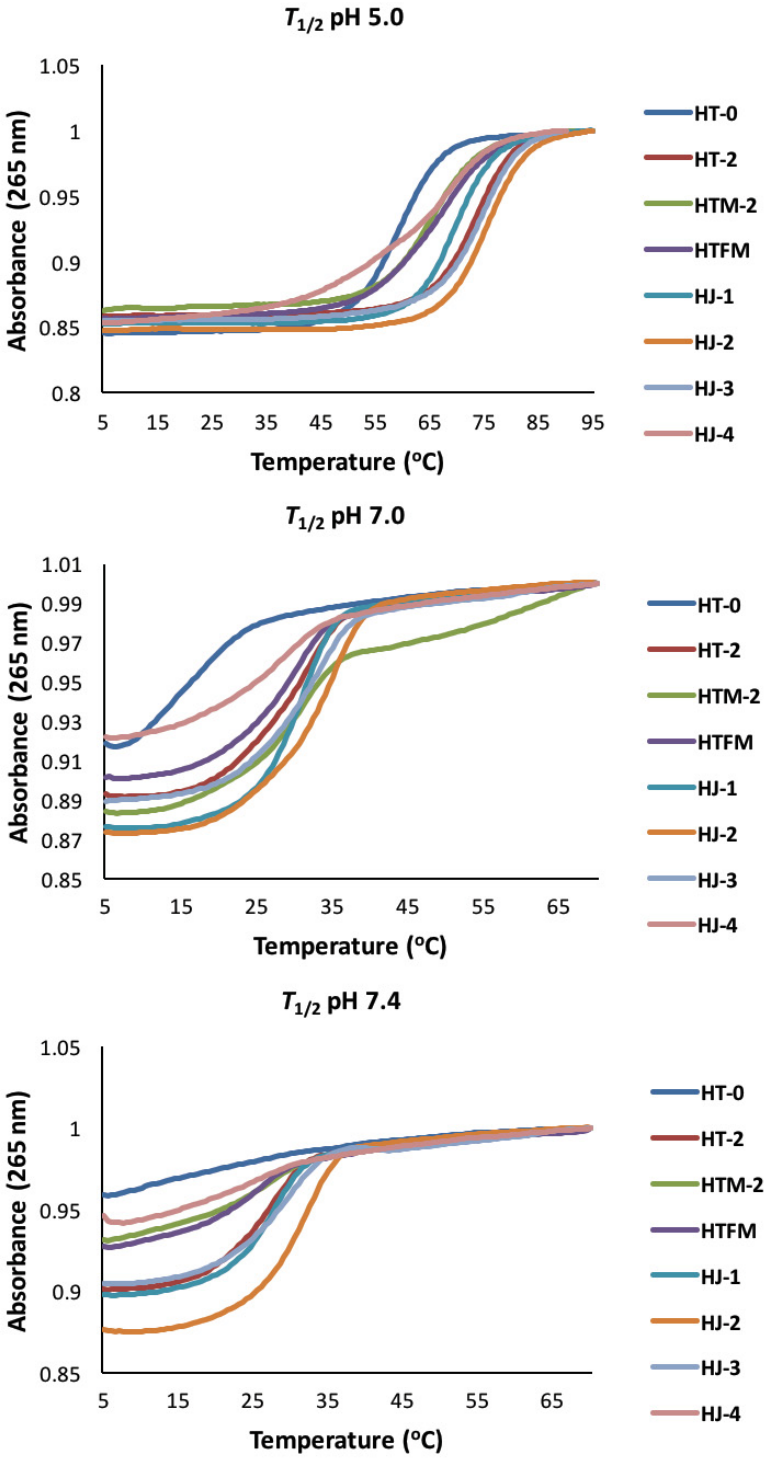


Figure 3.4: Thermal melting curves. Oligonucleotide concentration is 4 µM in 10 mM NaP_i, pH 5.0 to pH 7.4. $T_{1/2}$ spectra were recorded at 265 nm with heating ramped at 0.5 °C/min.

First, 5-Me-dC was incorporated in the human telomeric sequence at the same positions as previously reported for 2'F-araC: four incorporations in HTM-1, eight in HTM-2 and HTM-3, and twelve in HTM-4 (**Table 3.1**).[23] Recently, Xu *et al.* incorporated one, two (forming a base pair), and three consecutive 5-Me-dC units in a 24-nt human telomeric repeat.[25] This study revealed that one or two 5-Me-dC inserts stabilize DNA i-motif structures, while three consecutive modifications were destabilizing. Interestingly, in our studies, HTM-4, which contains four stretches of three consecutive 5-Me-dC modifications, was more stable than HT-0 at pH 5.0 ($\Delta T_{1/2}$ = +7.0 °C, **Table 3.1**) and at pH 7.0 ($\Delta T_{1/2}$ = +7.5 °C). The highest $T_{1/2}$ values at pH 7.0 were obtained for sequence HTM-2 ($T_{1/2}$ = 31.0 °C) with eight 5-Me-dC incorporations.

CD experiments at pH 5.0 confirmed i-motif formation, exhibiting characteristic positive and negative CD bands at approximately 285 nm and 255-260 nm, respectively (**Figure 3.5**). At pH 7.0, the strands having modified cytosine inserts retained the characteristic i-motif signature. By contrast, the control strand (HT-0) no longer showed the characteristic positive i-motif signature at pH 7.0, as expected.

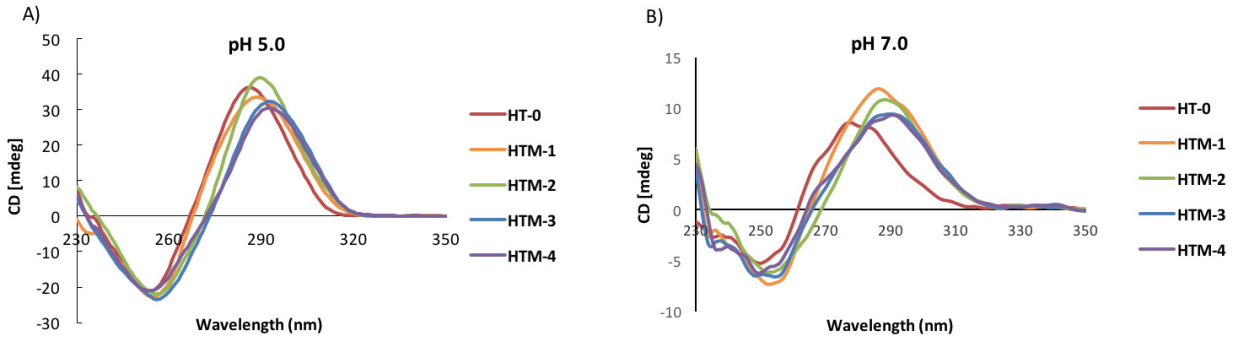


Figure 3.5: CD spectra recorded at 5 °C in 10 mM NaP_i. A) Oligonucleotide concentration is 50 µM in single strands. B) Oligonucleotide concentration is 20 µM in single strands.

This stabilization has been ascribed to the hydrophobicity and bulkiness of the 5-methyl group, which consequently leads to reduced flexibility of the methylated i-motif structures.[25] Moreover, Yang *et al*. reported that 5-methylcytosine stabilizes DNA i-motif conformations by increasing the base-pairing energies (BPEs) relative to cytosine.[38] From an electronic perspective, this increase in BPE is expected since the electron-donating methyl group stabilizes

the positive charge, strengthens the base-pairing interaction, and therefore stabilizes the methylated i-motif structure.

Based on these results, 5-Me-2'F-araC was incorporated at the same positions as in HTM-2 to provide sequence HTFM (**Table 3.1**). HTFM was more stable than HT-0 at all the studied pHs ($T_{1/2}$ = 70.4 °C *vs*. 59.2 °C, pH 5.0) and HTM-2 ($T_{1/2}$ = 64.8 °C). While HTFM exhibited a lower $T_{1/2}$ value relative to the 2'F-araC modified sequence (HT-2), it maintained significant stability at neutral pH ($T_{1/2}$ = 30.2 °C *vs.* 32.1 °C). At pH 7.0 and 7.4, HT-2, HTM-2, and HTFM were at least 15 °C more stable than the control sequence (**Table 3.1** and **Figure 3.4**).

Next, we investigated the effect of combining 5-Me-dC and 2'F-araC modifications in the same sequence (5-methyl and 2'-fluorine modifications in different nucleotides). It is interesting to note that while HJ-1 and HJ-2 each comprise four 5-Me-dC inserts and four 2'F-araC inserts, HJ-2 exhibited significantly higher stability at all studied values of pH. Based on the NMR studies described below, we hypothesize that the positional effect observed in these sequences arises from the nature of their base pairs (**fC**·**fC**$^+$ / **C**^{m}·**C**$^{m+}$ in HJ-1 *versus* **fC**·**C**$^{m+}$ in HJ-2, **Table 3.1**). HJ-3, containing four 2'F-araC and four 5-Me-2'F-araC inserts, exhibited very similar stability to HT-2 and HJ-2 at pH 5.0. At pH 7.0 and pH 7.4, the stability trend observed was HJ-2 > HJ-3 > HT-2. Sequence HJ-4 (containing only four 5-Me-2'F-araC modifications) exhibited the lowest $T_{1/2}$ value; nevertheless, HJ-4 was significantly more stable than the control sequence HT-0 in the pH 5.0-7.4 range.

3.4 Exploring the Effect of Chemical Modifications on the pK_a of Nucleosides and $pH_{1/2}$ of i-Motif Structures

Next we sought to determine the pK_a values of the newly synthesized nucleoside **6** and the $pH_{1/2}$ for some of the ON sequences in order to investigate the effect of combined modifications on the $pH_{1/2}$ of i-motif structures. The pK_a values of 5-Me-2'F-araC (pK_a = 4.2) and the other cytidine derivatives were determined by monitoring the change in absorbance at 280 nm as a function of pH (**Figure 3.6** and **Table 3.2**). As expected, 2'F-araC exhibited the lowest pK_a value (3.9) due to the electronegative nature of the fluorine atom, while 5-Me-dC exhibited the highest pK_a value (4.6) due to the presence of the electron-donating methyl group.

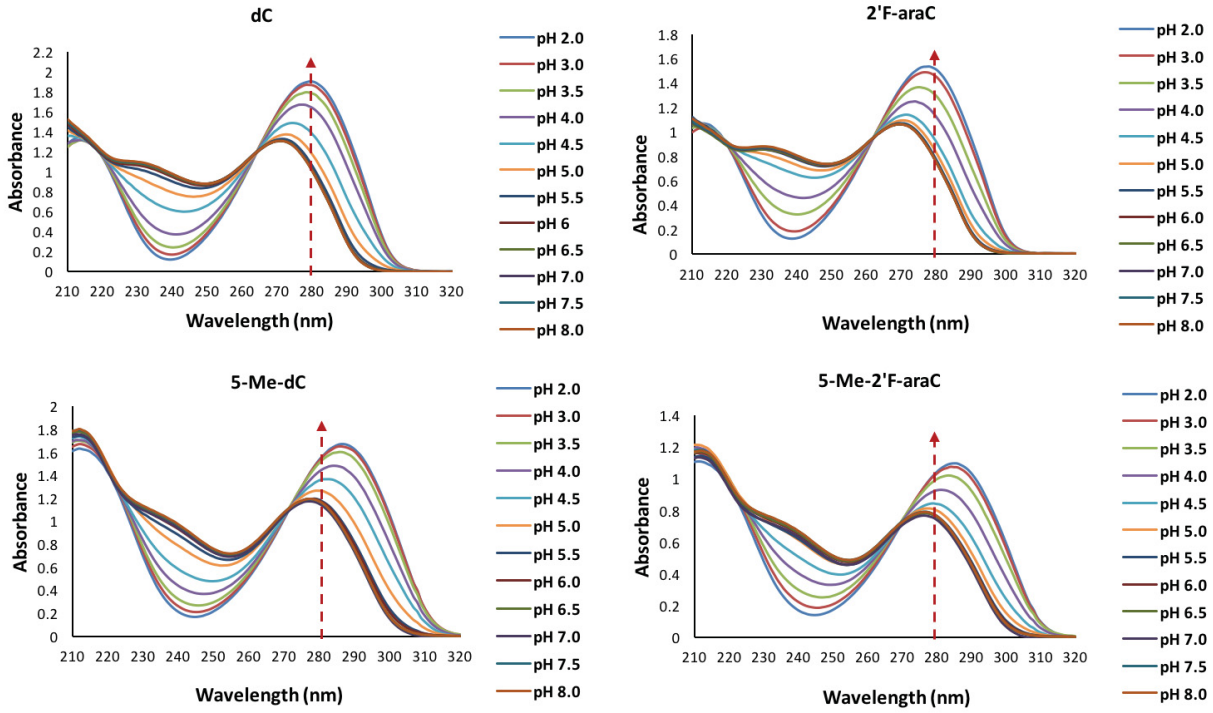


Figure 3.6: UV spectra for the free nucleosides as a function of pH.

We will refer to the pK_a of the i-motif structures as $pH_{1/2}$, defined as the pH at which 50% of the sequences are folded into i-motifs. CD spectra were performed from pH 4.0 to pH 9.0 (**Figure 3.7**). As the pH increases, the characteristic i-motif signature disappears and a positive band at 275 nm and a negative peak at 250-255 nm appear, indicating the formation of a single-stranded DNA random coil conformation. The $pH_{1/2}$ of the modified i-motif structures was determined by measuring the change in ellipticity at 284 nm as a function of pH. In order to determine the proportion of folded and unfolded states, the titration profiles were fit to a standard titration model (**Figure 3.8**). The $pH_{1/2}$ data of the i-motif structures generally correlated well with $T_{1/2}$ data (**Tables 3.1** and **3.2**). With the exception of HJ-1, all of the modified sequences exhibited a higher $pH_{1/2}$ value with respect to HT-0. Consistent with the $T_{1/2}$ data, the $pH_{1/2}$ value of HJ-2 was greater than that of HJ-1 (7.17 *versus* 6.53, **Table 3.2**). HJ-3 exhibited the highest $pH_{1/2}$ value (7.34), which very closely matches the physiological pH value (≈ 7.4). Therefore, whether 5-Me-dC and 2'F-araC (HJ-2, $pH_{1/2}$ 7.17) or 5-Me-2'F-araC (HTFM, $pH_{1/2}$ 7.34) were introduced, a significant increase in $pH_{1/2}$ was observed. This suggests that these modifications are well accommodated within the i-motif structure leading to favorable interactions translated in this increase in $pH_{1/2}$.

Table 3.2: Nucleoside pK_a and i-motif $pH_{1/2}$ values.

Nucleoside[a]	Nucleoside pK_a	Δ pK_a
2'-deoxycytidine	4.4 ± 0.02	-
2'-deoxy-2'-fluoroarabinocytidine	3.9 ± 0.05	-0.5
2'-deoxy-5-methylcytidine	4.6 ± 0.1	+0.2
2'-deoxy-2'-fluoro-5-methyl-arabinocytidine	4.2 ± 0.2	-0.2
Code[b]	**Oligonucleotide $pH_{1/2}$**	**$\Delta pH_{1/2}$**
HT-0	6.55 ± 0.3	-
HT-2	7.27 ± 0.4	+0.72
HTM-2	6.99 ± 0.3	+0.44
HTFM	7.13 ± 0.3	+0.58
HJ-1	6.53 ± 0.3	-0.02
HJ-2	7.17 ± 0.2	+0.62
HJ-3	7.34 ± 0.3	+0.79
HJ-4	6.70 ± 0.1	+0.15

[a]The pK_a values for the nucleosides were calculated from the plot of UV absorbance at 280 nm in 10 mM NaP_i buffer as a function of pH.
[b]The $pH_{1/2}$ values for i-motif structures were calculated from the plot of molar ellipticity at 284 nm *versus* pH in 10 mM NaP_i buffer at 5 °C. Oligonucleotide concentration is 50 µM in single strands. UV and CD data versus pH data were fit to a standard titration model involving a single protonation event using $CD_{obs} = CD_{high} + (CD_{low} - CD_{high})/(1 + 10^{pH-pKa})$. Errors were calculated according to the variance-covariance method.[39]

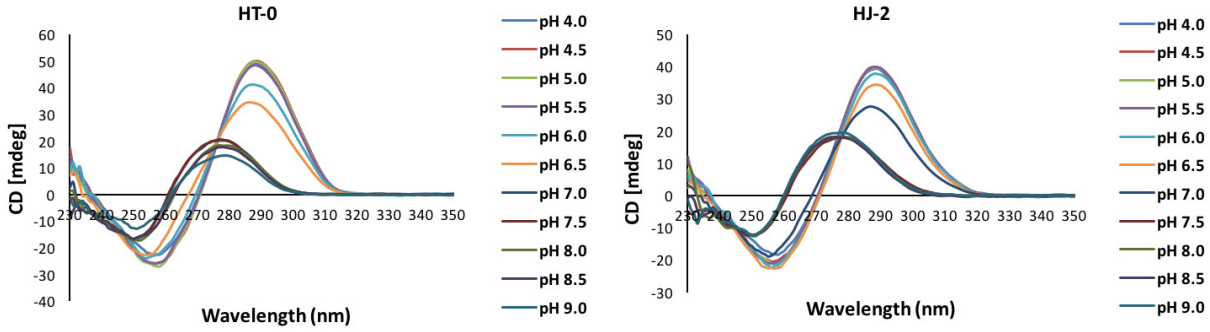


Figure 3.7: CD spectra for HT-0 (left) and HJ-2 (right) reported from pH 4.0 to 9.0.

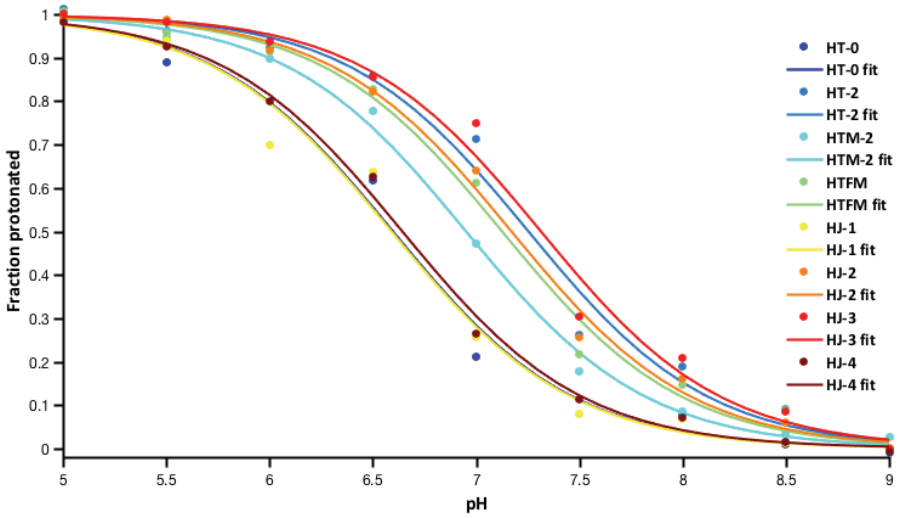


Figure 3.8: Plot of ellipticity at 284 nm as a function of pH at T = 5 °C. Experimental data are shown as colored filled circles, fits are shown as colored lines. Experimental data were fit to $CD_{obs} = CD_{high} + (CD_{low} - CD_{high})/(1 + 10^{pH-pKa})$.

3.5 Structural Determination *via* 2D NMR Spectroscopy

To obtain further insight on the stabilizing effect of the 2'F-araC / 5-Me-dC combination in i-motif structures, an NMR study of HJ-2 was undertaken in collaboration with Prof. Carlos González and Israel Serrano at the CSIC in Madrid. ^{1}H-NMR melting temperature experiments confirmed the high stability of HJ-2, with cytosine imino and amino protons observed at 15-16 and 8.5-10.5 ppm, respectively, at exceptionally high temperatures at neutral pH (up to 45 °C, **Figure 3.9A**). Based on the NOE assignments (**Figures 3.10** and **3.11**), the imino proton signals observed at 45 °C correspond to the **fC**·$\mathbf{C}^{m+}$ base pairs confirming that the stability of this specific structure, HJ-2, is due to 2'F-araC base-pairing with 5-Me-dC. The presence of 5-Me-dC and 2'F-araC substitutions improved the chemical shift dispersion in comparison with the unmodified sequence (**Figure 3.10**).

Two-dimensional NMR spectra of HJ-2 exhibited all the NOE patterns characteristic of i-motifs. Resonances of cytosine nucleobases were identified by their H5-H6 TOCSY cross-peaks and H5-amino NOEs (**Figure 3.11**). A similar pattern was observed for 5-Me-dC (*i.e.* Me5-H6 TOCSY cross-peak and Me5-amino NOEs). Nucleobase spin systems were connected with their own sugars through H6-H1' and H6-H2'/H2" NOE cross-peaks. 2'F-araC residues were distinguished from dC by the characteristic splitting in the H1' resonances of the fluorinated sugar, due to the strong ^{1}H-^{19}F coupling. Hemiprotonated base-pairs were identified by following

intra-nucleotide H6-H5/Me5 → intra-nucleotide H5/Me5-amino → inter-strand amino-imino → intra-nucleotide imino-amino → intra-nucleotide amino-H5/Me5 → intra-nucleotide H5/Me5-H6 pathways in the NOESY spectra (**Figure 3.10**). Four pathways were found connecting 5-Me-dC and 2'F-araC residues, and two connecting dC residues. This shows the formation of four **fC**·$\mathbf{C}^{m+}$ and two C·C^+ base pairs. H1'-H1' NOE cross-peaks along the minor grooves were found between 2'F-araCs on one side and between 5-Me-dCs on the other side (**Figure 3.9B**); therefore, all 2'F-araCs occupy one minor groove and base pair with the 5-Me-dCs. These connectivities, together with some weak sequential NOEs, allowed the unambiguous assignment of the residues in the central core of the i-motif.

Hence, the structure of HJ-2 is stabilized by four hemiprotonated **fC**·$\mathbf{C}^{m+}$ base pairs flanked by two C·C^+ base pairs (**Figure 3.9**). The experimental NOEs strongly indicate that the structure of HJ-2 is very similar to that of the native structure (HT-0), with both adopting a 5'E configuration.[4] Based on this similarity, a model structure of HJ-2 was built from the structure of the human telomeric C-rich strand reported by Phan and co-workers (PDB 1ELN).[4] After performing the appropriate mutations, the coordinates were submitted to a restrained molecular dynamics calculation with the AMBER package. A number of experimental distance constraints involving protons of the central i-motif core were included. The final model is shown in **Figure 3.9** and the chemical shifts are listed in **Table 3.3**.

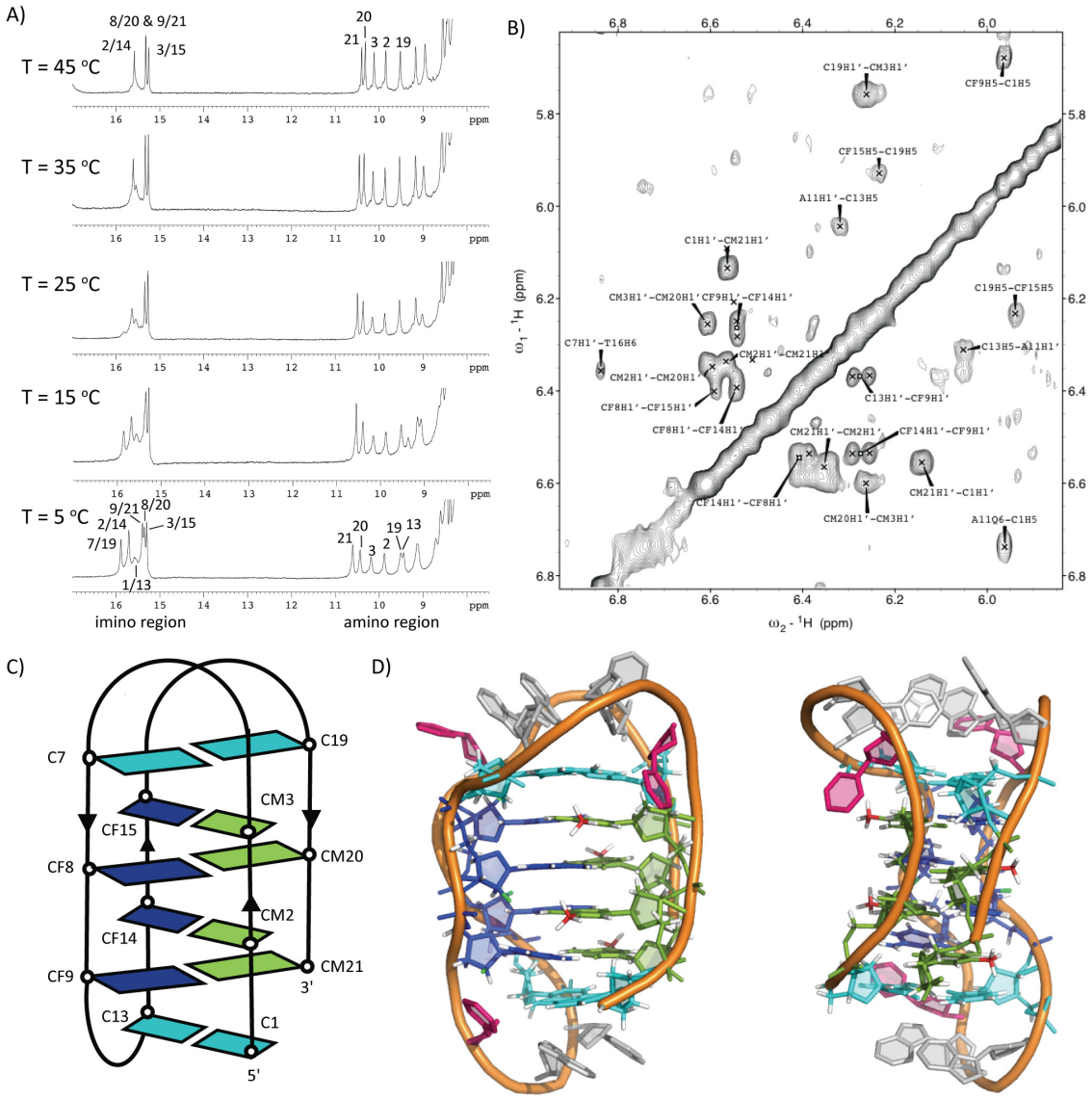


Figure 3.9: A) ^{1}H-NMR melting experiments of HJ-2 at pH 7.0. B) H1'-H1' region of the NOESY spectra of HJ-2 (mixing time 150 ms, T = 5 °C, pH 7.0). H1'-H1' connectivities are characteristic of i-motif structures. C) Schematic representation of HJ-2 structure with 5'E intercalation topology. D) Two views of the solution structure of HJ-2: the major groove side (left) and the 5-Me-dC minor groove side (right). Color code: dC in cyan; 5-Me-dC in green (5-Methyl group in red); 2'F-araC in blue (2'F atoms in green); dT in magenta; and dA in grey.

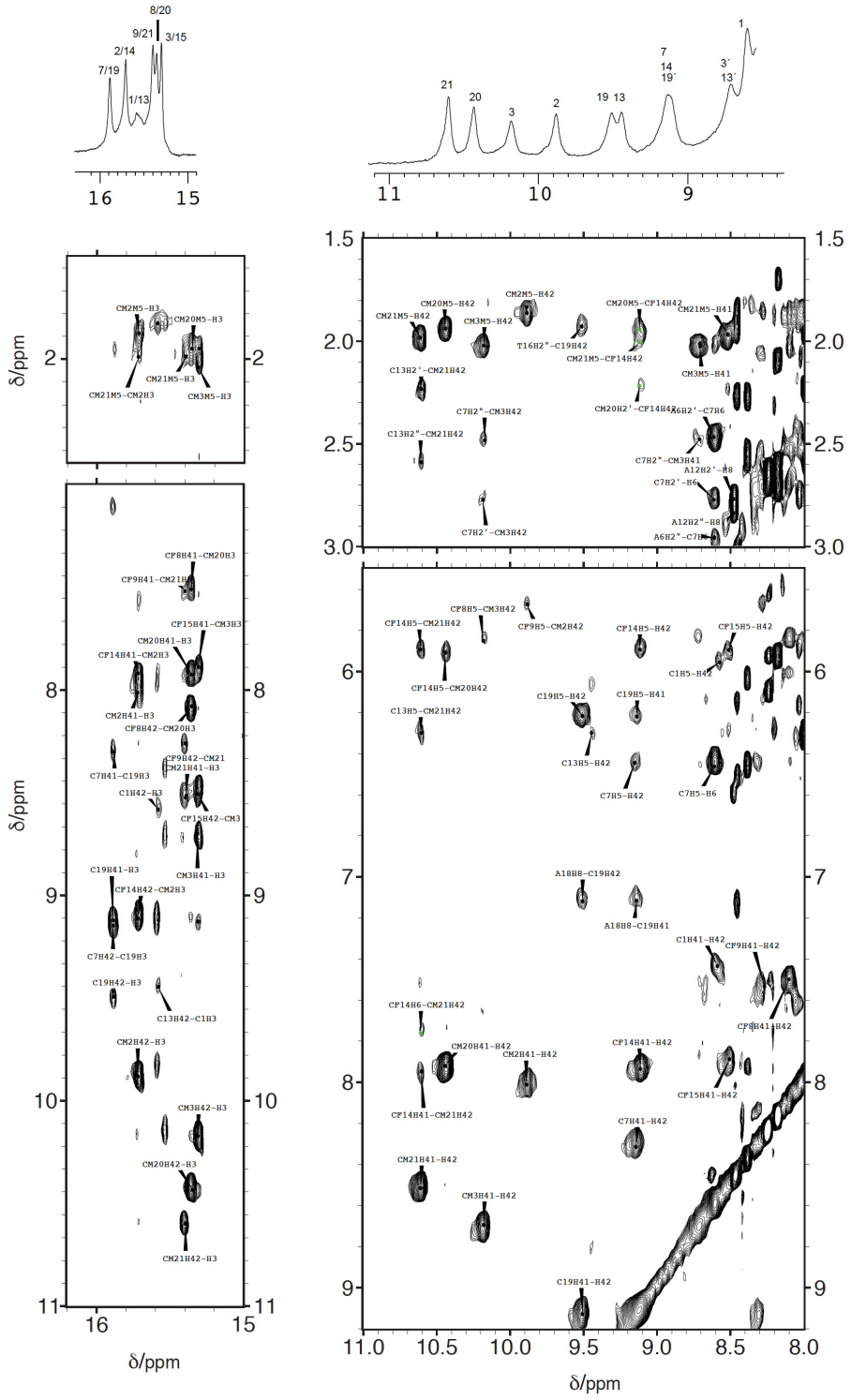


Figure 3.10: Regions of the NOESY spectra of HJ-2 (150 ms mixing time, T = 5 °C, pH 7.0). This figure shows the assigned characteristic cross peaks and illustrates the high quality of the NMR spectra with good dispersion of exchangeable protons. Chemical shifts are listed in **Table 3.3**.

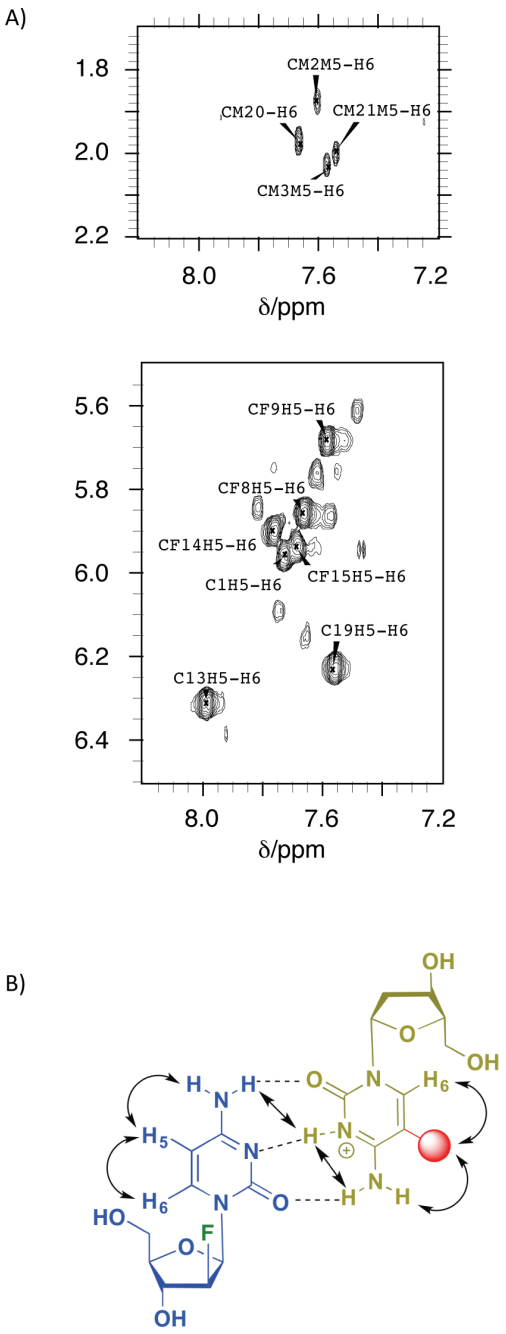


Figure 3.11: A) Regions of the TOCSY spectra of HJ-2 (mixing time 150 ms, T = 5 °C, pH 7.0) showing Me-H6 cross-peaks for 5-Me-dC (top) and H5-H6 cross-peaks for dC and 2'F-araC (bottom). B) Scheme representing 2'F-araC·5-Me-dC$^+$ base pair along with the inter-strand and intra-nucleotide NOEs observed in the NOESY spectra.

Table 3.3: Chemical shifts of HJ-2, pH 7.0, 10 mM sodium phosphate buffer, T = 5 °C.

	H1'	H2'	H2"	H3'	H5/H2/M	H6/8	H41	H42	H3
C1	6.14	2.05	2.54	4.79	5.96	7.73	7.46	7.46	15.58
CM2	6.34	2.04	2.54	4.63	1.88	7.61	8.02	8.02	15.71
CM3	6.26	2.31	2.81	4.84	2.03	7.58	8.71	8.71	15.30
C7	6.36	2.77	2.48	5.12	6.46	8.62	9.15	8.32	-
CF8	6.40	-	5.24	4.65	5.86	7.67	8.09	7.51	-
CF9	6.27	-	5.29	n.a.	5.68	7.59	8.27	7.52	-
C13	6.37	2.24	2.60	4.95	6.31	7.99	9.44	8.70	-
CF14	6.54	-	5.36	4.85	5.90	7.77	9.11	7.93	-
CF15	6.59	-	5.50	n.a.	5.94	7.69	8.51	7.89	-
C19	5.76	1.96	2.42	4.88	6.23	7.56	9.51	9.13	18.88
CM20	6.60	2.22	2.63	5.02	1.95	7.77	10.44	7.93	15.35
CM21	6.56	2.41	2.57	4.74	2.00	7.55	10.61	8.52	15.39

In spite of the overall similarity with the unmodified structure, some differences are observed: 2'F-araC sugars adopt a South conformation instead of the North conformation usually found in the native i-motif (**Figure 3.12**).

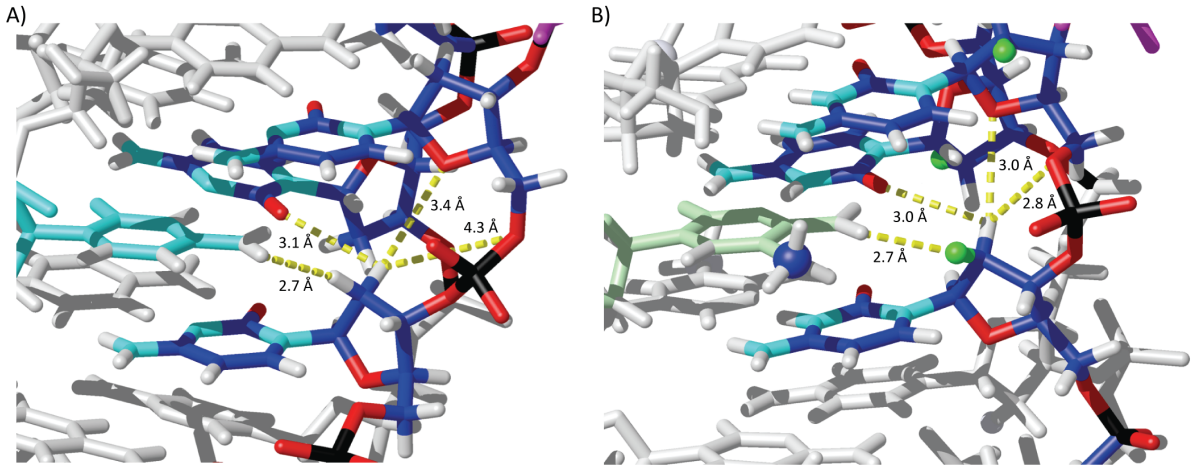


Figure 3.12: Details of sugar contacts in A) unmodified telomeric i-motif (PDB 1ELN) and B) 2'F-araC and 5-Me-dC substituted i-motif (HJ-2). Fluorine atom is shown in green and close distances corresponding to favorable electrostatic interactions are shown in yellow.

Similar to our recently reported observations in the tetrameric structure of d(TCCfCfCC),[23] the arabinose 2'-fluorine is well accommodated within the major groove of the i-motif. Also, HJ-2 exhibited very similar pattern of favourable electrostatic contacts that were previously obtained in the 2'F-araC tetrameric i-motif structure (**Figure 3.12**).[23] The

electronegative fluorine and the positively polarized H2' alleviate the unfavorable electrostatic interactions leading to closer sequential (H2'-O4', H2'-O5') and inter-strand distances (H2'-O2) compared to the unmodified structure.

3.6 Conclusion

In this chapter we were able to synthesize 5-Me-2'F-araC, a novel cytidine modification for use in i-motif studies, in very good yields. This nucleoside stabilizes i-motif structures when incorporated alone or in combination with other cytidine derivatives. Certain cytidine combinations afforded i-motif structures with $pH_{1/2}$ values close to physiological pH, highlighting the impact of base-pairing on i-motif stability. Based on the $T_{1/2}$ data obtained at pH 7.4, the following stabilization trend was observed: **fC**·$\mathbf{C}^{m+}$ > $\mathbf{fC}^{m}$·**fC**$^{+}$ > **fC**·**fC**$^{+}$ ≈ $\mathbf{C}^{m}$·$\mathbf{C}^{m+}$ > $\mathbf{fC}^{m}$·$\mathbf{fC}^{m+}$ ≈ $\mathbf{fC}^{m}$·C^{+} > C·C^{+}. By carefully selecting the positions of the various cytidine modifications studied here, it may be possible to fine-tune the pH and temperature-dependent responses of nanodevices constructed from i-motifs.[40] This study offers a library of chemically modified oligonucleotides that may be used to probe the biological significance of i-motifs that are prevalent in telomeric and promoter regions of the human genome. Furthermore, they may find use in the discovery and development of specific ligands or antibodies that target native i-motifs *in vivo*.

3.7 Experimental Data

3.7.1 General Considerations

Unless indicated otherwise, reactions were performed in oven-dried flasks with dry solvents under Argon. Solvents used were dried by MB-SPS 800 Auto (MBraun). Proton, fluorine, and phosphorous nuclear magnetic resonance spectra (^{1}H NMR, ^{19}F NMR, ^{31}P NMR) were acquired on Varian Mercury 400 MHz and Bruker AV500 spectrometers. Carbon magnetic resonance spectra (^{13}C NMR) were acquired on the Bruker AV500. Data were processed using TopSpin or MestReNova software. Chemical shifts (δ) are reported in parts per million (ppm), calibrated to the residual protium in the deuterated solvent. Coupling constants (J) are reported in Hz. Multiplicities are reported using the following abbreviations: *s* = singlet, *d* = doublet, *t* = triplet, *q* = quartet, *m* = multiplet (where the range of chemical shift is given). Analytical thin-plate chromatography was performed on pre-coated 200 μm layer thick silica gel TLA-R10011B-323 plates (Silicycle). Purifications by column chromatography were performed on

silica gel (40-63 μm, 230-400 mesh). High-resolution mass spectra (HRMS) were obtained on a mass spectrometer under electron spray ionization (ESI) conditions.

3.7.2 Experimental Procedure for All Synthesized Compounds

3.7.2.1 2-deoxy-2-fluoro-3,5-di-O-benzoyl-α-D-arabinofuranosyl bromide (2)

An oven-dried 250 mL round-bottom flask, equipped with a rubber septum and Teflon-coated stir bar, was charged with **1** (6.0 g, 12.92 mmol, 1 equiv), followed by CH_2Cl_2 (81 mL, 0.16 M) and 33 % HBr (9.6 mL) in acetic acid. The mixture was stirred under Argon at room temperature for 24 h. This led to the complete consumption of **1** as monitored by TLC (in CH_2Cl_2) using UV shadowing. The mixture was diluted with CH_2Cl_2 (50 mL), quenched with sat. $NaHCO_3$ (100 mL), and transferred to a separatory funnel. It was then washed with CH_2Cl_2 (100 mL x 3). The organic phase was dried over $MgSO_4$, filtered, and concentrated *in vacuo* to afford **2**, a brown oil. Yield: 95 %

R_f (CH_2Cl_2) = 0.73; **^{1}H NMR** (500 MHz, $CDCl_3$): δ = 8.10 - 7.44 (m, 10H, H-Bz), 6.66 (d, J_{HF} = 12.1 Hz, 1H, H-1), 5.62 (d, J_{HF} = 50.0 Hz, 1H, H-2), 5.57 (dd, J_{HF} = 22.0 Hz, J_{HH} = 3.3 Hz, 1H, H-3), 4.88 - 4.79 (m, 2H, H-4 and H-5a), 4.78 - 4.71 (m, 1H, H-5b); **^{19}F NMR** (500 MHz, $CDCl_3$): δ = -165.9 (ddd, J_{HF} = 52.0 Hz, J_{HF} = 23.5 Hz, J_{HF} = 12.5 Hz).

3.7.2.2 N^4-benzoyl-5-methylcytosine (4)

An oven-dried 250 mL round-bottom flask, equipped with a rubber septum and Teflon-coated stir bar, was charged with 5-methylcytosine **3** (1.8 g, 14.39 mmol, 1 equiv), followed by pyridine (43.3 mL, 0.33 M), and benzoylchloride (2.0 mL, 17.26 mmol, 1.2 equiv). The mixture was stirred under Argon at reflux for 24 h. This led to the complete consumption of **3** as monitored by TLC (3 % MeOH in CH_2Cl_2) using UV shadowing. The mixture was quenched with MeOH (100 mL) and concentrated *in vacuo*. The crude was purified by column chromatography (CH_2Cl_2 → 1 % MeOH in CH_2Cl_2) to afford **4**, a beige solid. Yield: 46 %

R_f (1% MeOH in CH_2Cl_2) = 0.36; **^{1}H NMR** (500 MHz, DMSO-d_6): δ = 8.20 (d, J_{HH} = 7.2 Hz, 2H, H-Bz), 7.66 (s, 1H, H-6), 7.58 (t, J_{HH} = 7.3 Hz, 1H, H-Bz), 7.49 (t, J_{HH} = 7.6 Hz, 2H, H-Bz), 1.99 (s, 3H); **^{13}C NMR** (125 MHz, DMSO-d_6): δ = 161.7, 142.3, 137.4, 132.8 (2C), 129.7 (3C), 128.7 (2C), 108.9, 13.1; **HRMS (ESI-)** m/z: calculated for $C_{12}H_{10}N_3O_2$ $[M - H]^-$ 228.0773; found 228.07761.

3.7.2.3 N^4-benzoyl-5-methyl-1-[2-deoxy-2-fluoro-3,5-di-*O*-benzoyl-β-arabinofuranosyl]cytosine (5)

An oven-dried 100 mL round-bottom flask, equipped with a rubber septum and Teflon-coated stir bar, was charged with **4** (800 mg, 3.49 mmol, 4 equiv), followed by HMDS (8.73 mL, 0.1 M*), and TMS-Cl (0.34 mL, 2.55 M*). The mixture was stirred under Argon at reflux for 8 h, at which point the solution became clear, signaling complete silylation of **4**. The reaction mixture was dried *in vacuo* and the flask purged with Argon immediately upon removal from vacuum to minimize its exposure to air. A solution of **2** (370 mg, 0.87 mmol, 1 equiv) dissolved in CCl_4 (4.85 mL, 0.18 M*) was syringed into the reaction flask. The mixture was stirred under Argon at reflux for 14 h. This led to the complete consumption of **2** as monitored by TLC (1 % MeOH in CH_2Cl_2) using UV shadowing. The mixture was quenched with sat. $NaHCO_3$ (50 mL), washed with brine (100 mL), and extracted with CH_2Cl_2 (50 mL x 3). The organic phase was dried over $MgSO_4$, filtered, and concentrated *in vacuo*. The crude product was purified by column chromatography (0.3 % MeOH in CH_2Cl_2) to separate the two anomers formed to afford **5**, a white solid. Yield: 41 %

Molarities of HMDS, TMS-Cl, and CCl_4 were calculated with respect to **2.*

R_f (1 % MeOH in CH_2Cl_2) = 0.24; **^{1}H NMR** (500 MHz, CD_2Cl_2): δ = 8.39 - 8.30 (m, 2H), 8.20 - 8.10 (m, 4H), 7.75 - 7.62 (m, 3H), 7.62 - 7.45 (m, 7H), 6.44 (dd, J_{HF} = 21.6 Hz, J_{HH} = 2.9 Hz, 1H, H-1'), 5.71 (dd, J_{HF} = 17.7 Hz, J_{HH} = 3.0 Hz, 1H, H-3'), 5.45 (dd, J_{HF} = 50.0 Hz, J_{HH} = 2.9 Hz, 1H, H-2'), 4.89 - 4.82 (m, 2H, H-5' and H-5''), 4.62 (q, J_{HH} = 3.9 Hz, 1H, H-4'), 2.01 (d, J_{HH} = 1.1 Hz, 3H, H-(5-Me)); **^{13}C NMR** (125 MHz, CD_2Cl_2): δ = 166.1, 165.2, 159.8 (2C), 147.6, 138.0, 137.0, 134.0, 133.4, 132.5, 128.7 - 128.1 (14C), 111.2, 92.7 (d, J_{CF} = 190.8 Hz), 85.1 (d, J_{CF} = 16.4 Hz), 81.4, 76.8 (d, J_{CF} = 30.6 Hz), 63.2, 13.2; **^{19}F NMR** (500 MHz, $CDCl_3$): δ = -200.91 - -201.10 (m); **HRMS (ESI+)** m/z: calculated for $C_{31}H_{27}FN_3O_7$ $[M + H]^+$ 572.1833; found 572.18304.

3.7.2.4 5-methyl-1-[2-deoxy-2-fluoro-β-D-arabinofuranosyl]cytosine (6)

A 100 mL round-bottom flask, equipped with a rubber septum and Teflon-coated stir bar, was charged with **5** (457 mg, 0.80 mmol, 1 equiv), followed by 2 M NH_3 in MeOH (18.0 mL, 0.04 M). The mixture was stirred under Argon at rt for 4 days. This led to the complete consumption of **5** as monitored by TLC (15 % MeOH in CH_2Cl_2) using UV shadowing. The

mixture was quenched with MeOH (50 mL) and concentrated *in vacuo*. The crude was purified by column chromatography (2 % MeOH in CH_2Cl_2 → 15 % MeOH in CH_2Cl_2) to afford **6**, a white solid. Yield: 95 %

R_f (15% MeOH in CH_2Cl_2) = 0.18; **^{1}H NMR** (500 MHz, MeOH-d_4): δ = 7.71 (s, 1H, H-6), 6.20 (dd, J_{HF} = 18.3 Hz, J_{HH} = 3.6 Hz, 1H, H-1'), 5.04 (ddd, J_{HF} = 52.1 Hz, J_{HH} = 3.7 Hz, J_{HH} = 2.1 Hz, 1H, H-2'), 4.31 (ddd, J_{HF} = 19.0 Hz, J_{HH} = 4.4 Hz, J_{HH} = 2.2 Hz, 1H, H-3'), 3.95 (q, J_{HH} = 4.3 Hz, 1H, H-4'), 3.85 (dd, J_{HH} = 12.2 Hz, J_{HH} = 3.3 Hz, 1H, H-5'), 3.77 (dd, J_{HH} = 12.1 Hz, J_{HH} = 5.6 Hz, 1H, H-5''), 1.98 (d, J_{HH} = 1.1 Hz, 3H, H-(5-Me)); **^{13}C NMR** (125 MHz, MeOH-d_4): δ =166.0, 156.5, 139.4, 102.2, 95.2 (d, J_{CF} = 190.0 Hz), 84.9 (d, J_{CF} = 16.8 Hz), 84.4, 73.9 (d, J_{CF} = 25.4 Hz), 60.7, 11.8; **^{19}F NMR** (500 MHz, MeOH-d_4): δ = -200.6 (dt, J_{HF} = 55.4 Hz, J_{HF} = 20.1 Hz); **HRMS (ESI+)** m/z: calculated for $C_{10}H_{15}FN_3O_4$ $[M + H]^+$ 260.1047; found 260.10381.

3.7.2.5 1-[2-deoxy-2-fluoro-3,5-di-*O*-benzoyl-β-D-arabinofuranosyl]thymine (8)

A 500 mL round-bottom flask, equipped with a rubber septum and Teflon-coated stir bar, was charged with **7** (9.31 g, 73.82 mmol, 5 equiv), followed by HMDS (147.5 mL, 0.1 M*), and TMS-Cl (5.8 mL, 2.53 M*). The mixture was stirred under Argon at reflux for 2 d, at which point the solution became clear, signaling complete silylation of **7**. The reaction mixture was dried *in vacuo* and the flask purged with Argon immediately upon removal from vacuum to minimize its exposure to air. A solution of **2** (6.22 g, 14.70 mmol, 1 equiv) dissolved in CCl_4 (147 mL, 0.1 M*) was syringed into the reaction flask. The mixture was stirred under Argon at reflux for 2 d. This led to the complete consumption of **2** as monitored by TLC (3% MeOH in CH_2Cl_2) using UV shadowing. The mixture was quenched with sat. $NaHCO_4$ (300 mL), washed with brine (300 mL) and extracted with CH_2Cl_2 (200 mL x 5). The organic phase was dried over $MgSO_4$, filtered, and concentrated *in vacuo*. The crude was purified by column chromatography (1 % MeOH in CH_2Cl_2 → 3 % MeOH in CH_2Cl_2) to afford **8**, a white solid. Yield: 84 %.

Molarities of HMDS, TMS-Cl, and CCl_4 were calculated with respect to **2.*

R_f (3% MeOH in CH_2Cl_2) = 0.18; **^{1}H NMR** (500 MHz, DMSO-d_6): δ = 11.54 (s, 1H, H-N^3), 8.08 - 8.06 (m, 4H, H-Bz), 7.74 - 7.66 (m, 2H, H-Bz), 7.60 - 7.52 (m, 4H, H-Bz), 7.43 (dd, J_{HH} = 2.5 Hz, J_{HH} = 1.4 Hz, 1H, H-6), 6.34 (dd, J_{HF} = 19.2 Hz, J_{HH} = 4.0 Hz, 1H, H-1'), 5.72 (ddd, J_{HF} = 20.3 Hz, J_{HH} = 4.8 Hz, J_{HH} = 1.9 Hz, 1H, H-3'), 5.56 (ddd, J_{HF} = 50.9 Hz, J_{HH} = 4.0

Hz, J_{HH} = 1.6 Hz, 1H, H-2'), 4.79 (dd, J_{HH} = 12.1 Hz, J_{HH} = 3.5 Hz, 1H, H-5'), 4.70 (dd, J_{HH} = 12.1 Hz, J_{HH} = 5.0 Hz, 1H, H-5''), 4.60 (q, J_{HH} = 4.7 Hz, 1H, H-4'), 1.64 (d, J_{HH} = 1.3 Hz, 3H, H-(5-Me)); **^{13}C NMR** (125 MHz, DMSO-d_6): δ =166.0, 165.3, 163.9, 150.5, 137.0, 134.4, 134.1, 130.1 - 129.1 (10C), 109.5, 93.8 (d, J_{CF} = 190.1 Hz), 83.4 (d, J_{CF} = 16.2 Hz), 78.7, 77.0 (d, J_{CF} = 29.2 Hz), 63.8, 12.4; **^{19}F NMR** (500 MHz, DMSO-d_6): δ = -199.2 (dt, J_{HF} = 53.5 Hz, J_{HF} = 22.1 Hz); **HRMS (ESI+)** m/z: calculated for $C_{24}H_{22}FN_2O_7$ $[M + H]^+$ 469.1411; found 469.14024.

3.7.2.6 5-methyl-1-[2-deoxy-2-fluoro-β-D-arabinofuranosyl]cytosine (6)

A 100 mL round-bottom flask, equipped with a rubber septum and Teflon-coated stir bar, was charged with **8** (500 mg, 1.07 mmol, 1 equiv), followed by 1,2,4-triazole (1.11 g, 16.01 mmol, 15 equiv), MeCN (21.3 mL, 0.05 M), and Et_3N (2.98 mL, 21.35 mmol, 20 equiv). The mixture was cooled to 0 °C in an ice bath. $POCl_3$ (0.4 mL, 4.27 mmol, 4 equiv) was then added dropwise, and the mixture was stirred under Argon at rt for 1 d. This led to the complete consumption of **8** as monitored by TLC (EtOAc) using UV shadowing to afford the intermediate **9**, which is fluorescent under UV. The mixture was concentrated *in vacuo*, and filtered through a silica plug with EtOAc (500 mL). The filtrate was again concentrated *in vacuo*. 33 % aq. NH_3 (60 mL, 0.2 M) was added. The mixture was stirred under Argon at rt for 4 h. This led to the complete consumption of **9** as monitored by TLC (EtOAc) using UV shadowing. The mixture was concentrated *in vacuo*. 2 M NH_3 in MeOH (60 mL, 0.2 M) was added, and the reaction was stirred under Argon at rt for 2 d. The crude was purified by column chromatography (CH_2Cl_2→ 15 % MeOH in CH_2Cl_2) to afford **6**, a white solid. Yield: 91 %

R_f (15% MeOH in CH_2Cl_2) = 0.20; **^{1}H NMR** (400 MHz, MeOH-d_4): δ = 7.74 (s, 1H , H-6), 6.21 (dd, J_{HF} = 18.1 Hz, J_{HH} = 3.7 Hz, 1H, H-1'), 5.04 (ddd J_{HF} = 52.1 Hz, J_{HH} = 3.6 Hz, J_{HH} = 2.2 Hz, 1H, H-2'), 4.32 (ddd, J_{HF} = 19.0 Hz, J_{HH} = 4.5 Hz, J_{HH} = 2.2 Hz, 1H, H-3'), 3.95 (q, J_{HH} = 4.3 Hz, 1H, H-4'), 3.85 (ddd, J_{HH} = 12.2 Hz, J_{HH} = 4.0 Hz, J_{HH} = 1.2 Hz, 1H, H-5'), 3.78 (dd, J_{HH} = 12.0 Hz, J_{HH} = 4.0 Hz, 1H, H-5''), 1.99 (d, J_{HH} = 1.1 Hz, 3H, H-(5-Me)); **^{13}C NMR** (125 MHz, MeOH-d_4): δ = 165.6, 156.4, 139.5, 102.7, 95.2 (d, J_{CF} = 190.1 Hz), 84.8 (d, J_{CF} = 16.8 Hz), 84.5, 73.7 (d, J_{CF} = 25.3 Hz), 60.6, 11.9; **^{19}F NMR** (500 MHz, MeOH-d_4): δ = -200.7 (dt, J_{HF} = 54.6 Hz, J_{HF} = 19.8 Hz); **HRMS (ESI+)** m/z: calculated for $C_{10}H_{15}FN_3O_4$ $[M + H]^+$ 260.1047; found 260.10416.

3.7.2.7 N^4-benzoyl-5-methyl-1-[2-deoxy-2-fluoro-β-D-arabinofuranosyl]cytosine (10)

A 100 mL round-bottom flask, equipped with a rubber septum and Teflon-coated stir bar, was charged with **6** (139 mg, 0.54 mmol, 1 equiv), followed by benzoic anhydride (146 mg, 0.64 mmol, 1.2 equiv), DMF (1 mL, 0.54 M), and pyridine (3 mL, 0.18 M). The mixture was stirred under Argon at rt for 2 d. The mixture was quenched with MeOH (50 mL) and concentrated *in vacuo*. The crude was purified by column chromatography (1 % MeOH in CH_2Cl_2 → 3 % MeOH in CH_2Cl_2) to afford **10**, a white solid. Yield: 42 %

R_f (10% MeOH in CH_2Cl_2) = 0.33; **^{1}H NMR** (500 MHz, MeOH-d_4): δ = 8.24 (d, J_{HH} = 7.7 Hz, 2H), 8.09 - 7.94 (m, 1H, H-6), 7.60 - 7.53 (m, 1H), 7.49 - 7.45 (m, 2H), 6.24 (dd, J_{HF} = 16.7 Hz, J_{HH} = 3.9 Hz, 1H, H-1'), 5.11 (ddd, J_{HF} = 52.2 Hz, J_{HH} = 4.0 Hz, J_{HH} = 2.5 Hz, 1H, H-2'), 4.38 (ddd, J_{HF} = 19.1 Hz, J_{HH} = 4.7 Hz, J_{HH} = 2.6 Hz, 1H, H-3'), 3.99 (q, J_{HH} = 4.6 Hz, 1H, H-4'), 3.89 (ddd, J_{HH} = 12.2 Hz, J_{HH} = 3.8 Hz, J_{HH} = 1.3 Hz, 1H, H-5'), 3.80 (dd, J_{HH} = 12.2 Hz, J_{HH} = 5.2 Hz, 1H, H-5''), 2.13 (d, J_{HH} = 1.1 Hz, 3H, H-(5-Me)); **^{13}C NMR** (125 MHz, acetone-d_6): δ = 179.2, 160.3, 147.5, 139.1, 137.3, 132.4, 129.7 (2C), 128.1 (2C), 110.0, 95.7 (d, J_{CF} = 190.7 Hz), 84.4, 83.9 (d, J_{CF} = 16.8 Hz), 73.4 (d, J_{CF} = 24.4 Hz), 60.5, 12.6; **^{19}F NMR** (500 MHz, MeOH-d_4): δ = -200.6 (dt, J_{HF} = 55.5 Hz, J_{HF} = 19.2 Hz); **HRMS (ESI+)** m/z: calculated for $C_{17}H_{19}FN_3O_5$ $[M + H]^+$ 364.1309; found 364.12997.

3.7.2.8 N^4-benzoyl-5-methyl-1-[2-deoxy-2-fluoro-5-(4,4'-dimethoxytrityl)-β-D-arabinofuranosyl]cytosine (11)

A 50 mL round-bottom flask, equipped with a rubber septum and Teflon-coated stir bar, was charged with **10** (88 mg, 0.24 mmol, 1 equiv), followed by 4-dimethylaminopyridine (3 mg, 0.024 mmol, 0.1 equiv), pyridine (1.94 mL, 0.125 M), and 4,4'-dimethoxytritylchloride (123 mg, 0.36 mmol, 1.5 equiv). The mixture was stirred under Argon at rt for 7 h. The crude was washed with CH_2Cl_2 (1 L) on silica that was neutralized with 3 % Et_3N, and purified by column chromatography (0.5 % MeOH in CH_2Cl_2 → 1 % MeOH in CH_2Cl_2) to afford **11**, a pale yellow solid. Yield: 62 %

R_f (10 % MeOH in CH_2Cl_2) = 0.74; **^{1}H NMR** (500 MHz, MeOH-d_4): δ = 8.21 (d, J_{HH} = 7.6 Hz, 2H), 7.80 (s, 1H, H-6), 7.56 - 7.43 (m, 5H), 7.38 - 7.22 (m, 7H), 6.89 - 6.86 (m, 4H), 6.27 (dd, J_{HF} = 16.6 Hz, J_{HH} = 4.1 Hz, 1H, H-1'), 5.11 (ddd, J_{HF} = 52.3 Hz, J_{HH} = 4.1 Hz, J_{HH} =

2.4 Hz, 1H, H-2'), 4.50 (ddd, J_{HF} = 19.8 Hz, J_{HH} = 4.8 Hz, J_{HH} = 2.5 Hz 1H, H-3'), 4.09 (q, J_{HH} = 4.3 Hz, 1H, H-4'), 3.77 (d, J_{HH} = 0.7 Hz, 6H, H-DMTr(-OMe)), 3.49 (dd, J_{HH} = 10.6 Hz, J_{HH} = 5.0 Hz, 1H, H-5'), 3.40 (dd, J_{HH} = 10.5 Hz, J_{HH} = 4.7 Hz, 1H, H-5''), 1.80 (d, J_{HH} = 1.1 Hz, 3H, H-(5-Me)); **^{13}C NMR** (125 MHz, MeOH-d_4): δ = 158.9 (2C), 144.8 (3C), 135.6, 132.3 - 126.6 (18C), 112.8 (4C), 110.4, 95.6 (d, J_{CF} = 191.0 Hz), 86.3, 84.5 (d, J_{CF} = 17.0 Hz), 83.3, 74.3 (d, J_{CF} = 25.4 Hz), 62.2, 54.3 (2C), 12.4; **^{19}F NMR** (500 MHz, MeOH-d_4): δ = -199.0 (dt, J_{HF} = 55.0 Hz, J_{HF} = 19.9 Hz); **HRMS (ESI+)** m/z: calculated for $C_{38}H_{37}FN_3O_7$ $[M + H]^+$ 666.2616; found 666.26066.

3.7.2.9 N^4-benzoyl-5-methyl-1-[3-(2-cyanoethoxy(diisopropylamino)-phosphinyl)-2-deoxy-2-fluoro-5-(4,4'- dimethoxytrityl)-β-D-arabinofuranosyl]cytosine (12)

A 50 mL round-bottom flask, equipped with a rubber septum and Teflon-coated stir bar, was charged with **11** (97 mg, 0.15 mmol, 1 equiv), followed by THF (1.46 mL, 0.1 M), N,N-diisopropylethylamine (0.13 mL, 0.73 mmol, 5 equiv), and chloro(2-cyanoethoxy)(diisopropylamino)phosphine. The mixture was stirred under Argon at rt for 6 h. The mixture was directly loaded on silica that was neutralized with 3 % Et_3N for purification by column chromatography (10 % EtOAc in hexanes → 40 % EtOAc in hexanes). The product was co-evaporated with acetone to afford **12**, a yellow solid. Yield: 85 %. The mixture of two diastereomers at phosphorous led to complex ^{1}H and ^{13}C NMR spectra.

R_f (1:1 EtOAc:hexanes) = 0.50; **^{19}F NMR** (500 MHz, MeCN-d_3): δ = -197.8 - -198.2 (m); **^{31}P NMR** (500 MHz, MeCN-d_3): δ = 151.61 (d, J_{PF} = 7.8 Hz), 150.12 (d, J_{PF} = 9.4 Hz); **HRMS (ESI+)** m/z: calculated for $C_{47}H_{53}FN_5O_8PNa$ $[M + Na]^+$ 888.3513; found 888.3508.

3.7.3 Oligonucleotide Synthesis and Purification

Oligonucleotide synthesis was performed at 1 μmol scale on the Unylinker CPG solid support (ChemGenes) on an ABI 3400 DNA Synthesizer (Applied Biosystems). Phosphoramidites thymidine (dT), deoxyadenosine (*N*-Bz) (dA), deoxycytidine (*N*-acetyl) (dC), and 5-methyl-2'-deoxycytidine (5-Me-dC) were used at 0.1 M concentration in acetonitrile and coupled for 200 s. Deoxyguanosine (*N*-ibu) (dG) was coupled for 300 s, 2'F-araC and 2'F-araG were used at 0.15 M concentration and coupled for 600 s and 900 s, respectively. 5-Me-2'F-araC was coupled for 1200 s. After the completion of synthesis, the CPG was transferred to a 1.5 mL

screw-cap eppendorf where aqueous NH_4OH (1 mL) was added and placed on a shaker for 48 hours at room temperature. The solution was centrifuged and then decanted from the cleaved CPG. Samples were vented for 1 h, chilled on dry ice, and evaporated to dryness, then purified by ion exchange HPLC on an Agilent 1200 Series HPLC (Agilent Technologies) using a Protein-Pak DEAE 5PW column (7.5 x 75 mm). The buffer system consisted of water (solution A) and 1 M aqueous lithium perchlorate (solution B) at a flow rate of 1 mL/min. The gradient was 0-60% solution B over 42 minutes at 60 °C. Samples were desalted using NAP-25 desalting columns according to manufacturer's procedure. The extinction coefficients of 2'F-araC and 5-Me-2'F-araC were approximated to that of the unmodified sequence using the Oligo Analyzer tool provided by IDT. Unless otherwise indicated, MilliQ water and autoclaved consumables were used. Masses were verified by ESI-MS (**Table 3.4**).

Table 3.4: MS characterization of the studied sequences.

Code	Sequence (5'-3')	Calculated mass (g/mole)	Mass found (g/mole) (-ve mode)
HT-0	CCC TAA CCC TAA CCC TAA CCC	6197.0846	6197.1250
HT-2	CfCfC TAA CfCfC TAA CfCfC TAA CfCfC	6341.0086	6341.0313
HTM-1	C**C^m**C TAA C**C^m**C TAA C**C^m**C TAA C**C^m**C	6253.1470	6253.0938
HTM-2	C**C^{m}C^m** TAA C**C^{m}C^m** TAA C**C^{m}C^m** TAA C**C^{m}C^m**	6309.2094	6309.1875
HTM-3	**C^{m}C^m**C TAA **C^{m}C^m**C TAA **C^{m}C^m**C TAA **C^{m}C^m**C	6309.2094	6309.1563
HTM-4	**C^{m}C^{m}C^m** TAA **C^{m}C^{m}C^m** TAA **C^{m}C^{m}C^m** TAA **C^{m}C^{m}C^m**	6365.2718	6365.2188
HTFM	C**fC^{m}fC^m** TAA C**fC^{m}fC^m** TAA C**fC^{m}fC^m** TAA C**fC^{m}fC^m**	6453.1342	6452.9688
HJ-1	CfCC^m TAA CC^{m}fC TAA CfCC^m TAA CC^{m}fC	6325.1090	6324.8125
HJ-2	C**C^{m}C^m** TAA CfCfC TAA CfCfC TAA C**C^{m}C^m**	6325.1090	6325.0938
HJ-3	C**fC^{m}fC^m** TAA CfCfC TAA CfCfC TAA C**fC^{m}fC^m**	6397.0714	6396.8125
HJ-4	C**fC^{m}fC^m** TAA CCC TAA CCC TAA C**fC^{m}fC^m**	6325.1094	6324.8125

(fC): 2'F-araC, (**C^m**): 5-Me-dC, and **fC^m**: 5-Me-2'F-araC

3.7.4 UV Melting Studies

UV thermal denaturation data were obtained on the Varian CARY 300 UV-Vis spectrophotometer (Agilent Technologies) equipped with a Peltier temperature controller. The concentration of oligonucleotides used was 4 µM prepared in 10 mM sodium phosphate buffer (pH 5.0, pH 7.0, and pH 7.4). Concentrations were determined after quantitating the samples by

UV absorbance at λ = 260 nm. Samples were heated at 90 °C for 15 minutes, slowly cooled to room temperature, then stored at 5 °C for at least an overnight before measurements were performed. Denaturation curves were acquired at 265 nm at a ramp of 0.5 °C/min. Samples were kept under a nitrogen flow at temperatures below 12 °C. The optimal melting temperature ($T_{1/2}$) values were calculated using the first derivative method.

3.7.5 Circular Dichroism Studies

CD studies were performed at 5 °C on the JASCO J-810 spectropolarimeter (JASCO Analytical Instruments) using a 1 mm path length cuvette. Temperature was maintained using the Peltier unit within the instrument. Spectra were recorded from 350-230 nm at a scan rate of 100 nm/min and a response time of 2.0 s with three acquisitions recorded for each spectrum. The spectra were normalized by subtraction of the background scan with the corresponding buffer. Data were smoothed using the means-movement function in the JASCO graphing software. The concentration of oligonucleotides used was 50 μM prepared in 10 mM sodium phosphate buffer (pH 5.0 and pH 7.0).

3.7.6 Nuclear Magnetic Resonance

Samples for NMR experiments were dissolved in 9:1 H_2O/D_2O (10 mM sodium phosphate buffer). All NMR spectra were acquired using Bruker Avance spectrometers operating at 600 and 800 MHz, equipped with cryoprobes, and processed with TOPSPIN software. NOESY[41] spectra were acquired with mixing times of 150 and 250 ms. TOCSY[42] spectra were recorded with the standard MLEV-17 spin-lock sequence and a mixing time of 80 ms. Water suppression was achieved by including a WATERGATE[43] module in the pulse sequence prior to acquisition. The spectral analysis program SPARKY was used for semiautomatic assignment of the NOESY cross-peaks and quantitative evaluation of the NOE intensities.

3.7.7 Structural Determination

Restrained molecular dynamics calculations were carried out with the SANDER module of the package AMBER 12.0.[44] Coordinates from the human telomeric i-motif (pdb code 1ELN) were utilized as starting points for the AMBER refinement, consisting of an annealing protocol in vacuo, followed by trajectories of 500 ps each in which explicit solvent molecules were included. The Particle Mesh Ewald method was used to evaluate long-range electrostatic

interactions. The specific protocols for these calculations have been described elsewhere.[45] The BSC1 force field was used to describe the DNA,[46] and the TIP3P model was used to simulate water molecules. Suitable parameters for the 2'F-araC derivatives were extracted from Noy *et al*.[47] Analysis of the representative structures was carried out with MOLMOL program.[48]

3.8 References

1. Gehring, K.; Leroy, J. L.; Gueron, M., A Tetrameric DNA-Structure with Protonated Cytosine.Cytosine Base-Pairs. Nature 1993, 363 (6429), 561-565.
2. Day, H. A.; Pavlou, P.; Waller, Z. A., i-Motif DNA: structure, stability and targeting with ligands. Bioorg. Med. Chem. 2014, 22 (16), 4407-18.
3. Brooks, T. A.; Kendrick, S.; Hurley, L., Making sense of G-quadruplex and i-motif functions in oncogene promoters. FEBS J. 2010, 277 (17), 3459-69.
4. Phan, A. T.; Gueron, 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.
5. 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.
6. 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.
7. 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.
8. 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.
9. 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. J. Am. Chem. Soc. 2014, 136 (11), 4172-4185.
10. Cui, Y. X.; Kong, D. M.; Ghimire, C.; Xu, C. X.; Mao, H. B., Mutually Exclusive Formation of G-Quadruplex and i-Motif Is a General Phenomenon Governed by Steric Hindrance in Duplex DNA. Biochemistry 2016, 55 (15), 2291-2299.
11. Benabou, S.; Aviñó, A.; Eritja, R.; González, C.; Gargallo, R., Fundamental aspects of the nucleic acid i-motif structures. Rsc Adv. 2014, 4 (51), 26956.
12. Robidoux, S.; Klinck, R.; Gehring, K.; Damha, M. J., Association of branched oligonucleotides into the i-motif. J. Biomol. Struct. Dyn. 1997, 15 (3), 517-27.

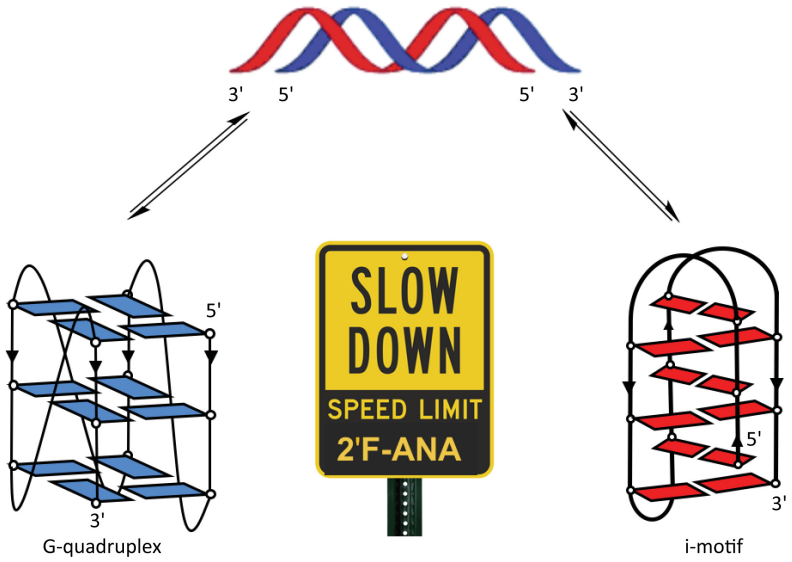
13. Li, X.; Peng, Y.; Ren, J.; Qu, X., Carboxyl-modified single-walled carbon nanotubes selectively induce human telomeric i-motif formation. Proc. Natl. Acad. Sci. USA 2006, 103 (52), 19658-63.
14. Chen, X.; Zhou, X.; Han, T.; Wu, J.; Zhang, J.; Guo, S., Stabilization and induction of oligonucleotide i-motif structure via graphene quantum dots. ACS Nano 2013, 7 (1), 531-7.
15. Kendrick, S.; Kang, H. J.; Alam, M. P.; Madathil, M. M.; Agrawal, P.; Gokhale, V.; Yang, D. Z.; 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.
16. 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), 529-35.
17. 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.
18. Snoussi, K.; Nonin-Lecomte, S.; Leroy, J. L., The RNA i-motif. J. Mol. Biol. 2001, 309 (1), 139-153.
19. Kumar, N.; Nielsen, J. T.; Maiti, S.; Petersen, M., i-Motif formation with locked nucleic acid (LNA). Angew. Chem. 2007, 46 (48), 9220-2.
20. 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), 752-5.
21. Perlíková, P.; Karlsen, K. K.; Pedersen, E. B.; Wengel, J., Unlocked Nucleic Acids with a Pyrene-Modified Uracil: Synthesis, Hybridization Studies, Fluorescent Properties and i-Motif Stability. Chembiochem 2014, 15 (1), 146-156.
22. Fenna, C. P.; Wilkinson, V. J.; Arnold, J. R.; Cosstick, R.; Fisher, J., The effect of 2'-fluorine substitutions on DNA i-motif conformation and stability. Chem. Commun. 2008, (30), 3567-9.
23. Abou Assi, H.; Harkness, R. W. V.; Martin-Pintado, N.; Wilds, C. J.; Campos-Olivas, R.; Mittermaier, A. K.; Gonzalez, C.; Damha, M. J., Stabilization of i-motif structures by 2'-beta-fluorination of DNA. Nucleic Acids Res. 2016, 44 (11), 4998-5009.
24. Lannes, L.; Halder, S.; Krishnan, Y.; Schwalbe, H., Tuning the pH Response of i-Motif DNA Oligonucleotides. Chembiochem 2015, 16 (11), 1647-1656.
25. Xu, B. C.; Devi, G.; Shao, F. W., Regulation of telomeric i-motif stability by 5-methylcytosine and 5-hydroxymethylcytosine modification. Org. Biomol. Chem. 2015, 13 (20), 5646-5651.
26. 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.

27. Brooks, T. A.; Hurley, L. H., Targeting MYC Expression through G-Quadruplexes. Genes Cancer 2010, 1 (6), 641-649.
28. 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), e11647.
29. Jin, S. G.; Kadam, S.; Pfeifer, G. P., Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. Nucleic Acids Res. 2010, 38 (11), e125.
30. Yildirim, O.; Li, R.; Hung, J. H.; Chen, P. B.; Dong, X.; Ee, L. S.; Weng, Z.; Rando, O. J.; Fazzio, T. G., Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. Cell 2011, 147 (7), 1498-510.
31. Kraus, T. F.; Globisch, D.; Wagner, M.; Eigenbrod, S.; Widmann, D.; Munzel, M.; Muller, M.; Pfaffeneder, T.; Hackner, B.; Feiden, W.; Schuller, U.; Carell, T.; Kretzschmar, H. A., Low values of 5-hydroxymethylcytosine (5hmC), the "sixth base," are associated with anaplasia in human brain tumors. Int. J. Cancer 2012, 131 (7), 1577-90.
32. Bird, A., DNA methylation patterns and epigenetic memory. Genes & Dev. 2002, 16 (1), 6-21.
33. Bird, A., Perceptions of epigenetics. Nature 2007, 447 (7143), 396-8.
34. Gonzalo, S.; Jaco, I.; Fraga, M. F.; Chen, T. P.; Li, E.; Esteller, M.; Blasco, M. A., DNA methyltransferases control telomere length and telomere recombination in mammalian cells. Nat. Cell Biol. 2006, 8 (4), 416-U66.
35. Ng, L. J.; Cropley, J. E.; Pickett, H. A.; Reddel, R. R.; Suter, C. M., Telomerase activity is associated with an increase in DNA methylation at the proximal subtelomere and a reduction in telomeric transcription. Nucleic Acids Res. 2009, 37 (4), 1152-1159.
36. Elzagheid, M. I.; Viazovkina, E.; Damha, M. J., Synthesis of protected 2'-deoxy-2'-fluoro-beta-D-arabinonucleosides. Current protocols in nucleic acid chemistry / edited by Serge L. Beaucage ... [et al.] 2002, Chapter 1, Unit 17.
37. Wilds, C. J.; Damha, M. J., 2'-Deoxy-2'-fluoro-beta-D-arabinonucleosides and oligonucleotides (2'F-ANA): synthesis and physicochemical studies. Nucleic Acids Res. 2000, 28 (18), 3625-35.
38. 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.
39. Tellinghuisen, J., Statistical error propagation. J. Phys. Chem. A 2001, 105 (15), 3917-3921.
40. Dong, Y. C.; Yang, Z. Q.; Liu, D. S., DNA Nanotechnology Based on i-Motif Structures. Acc. Chem. Res. 2014, 47 (6), 1853-1860.
41. Kumar, A.; Ernst, R. R.; Wuthrich, 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.
42. Bax, A.; Davis, D. G., Mlev-17-Based Two-Dimensional Homonuclear Magnetization Transfer Spectroscopy. J. Magn. Reson. 1985, 65 (2), 355-360.
43. Piotto, M.; Saudek, V.; Sklenar, V., Gradient-Tailored Excitation for Single-Quantum Nmr-Spectroscopy of Aqueous-Solutions. J. Biomol. NMR 1992, 2 (6), 661-665.
44. Case, D. A.; Darden, T. A.; Cheatham, T. E. III.; et al. AMBER 12, University of California, San Francisco, 2012.
45. Soliva, R.; Monaco, V.; Gomez-Pinto, I.; Meeuwenoord, N. J.; Van der Marel, G. A.; Van Boom, J. H.; Gonzalez, C.; Orozco, M., Solution structure of a DNA duplex with a chiral alkyl phosphonate moiety. Nucleic Acids Res. 2001, 29 (14), 2973-2985.
46. Ivani, I.; Dans, P. D.; Noy, A.; Perez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goni, R.; Balaceanu, A.; Portella, G.; Battistini, F.; Gelpi, J. L.; Gonzalez, C.; Vendruscolo, M.; Laughton, C. A.; Harris, S. A.; Case, D. A.; Orozco, M., Parmbsc1: a refined force field for DNA simulations. Nat. Methods 2016, 13 (1), 55-8.
47. Noy, A.; Luque, F. J.; Orozco, M., Theoretical analysis of antisense duplexes: determinants of the RNase H susceptibility. J. Am. Chem. Soc. 2008, 130 (11), 3486-96.
48. Koradi, R.; Billeter, M.; Wuthrich, K., MOLMOL: A program for display and analysis of macromolecular structures. J. Mol. Graph. 1996, 14 (1), 51.

Chapter 4

2'-Fluoroarabinonucleic Acid Modification Traps G-Quadruplex and i-Motif Structures in Human Telomeric DNA



The majority of this chapter is reproduced from: "2'-Fluoroarabinonucleic Acid Modification Traps G-Quadruplex and i-Motif Structures in Human Telomeric DNA", **Hala Abou Assi**, Roberto El-Khoury, Carlos González, and Masad J. Damha, *Nucleic Acids Research*, **2017**, *45*, 11535-11546.

"Discipline and freedom are not mutually exclusive but mutually dependent because otherwise, you'd sink into chaos."
–Paulo Coelho

4.1 Introduction

Human telomeric DNA is composed of tandem guanine-rich repeats $(TTAGGG)_n$ along with complementary cytosine-rich repeats $(CCCTAA)_n$.[1] The G-rich sequence forms G4 structures consisting of square-planar stacked guanine tetrads stabilized by Hoogsteen hydrogen bonds and mono- or divalent cations (**Figure 4.1**).[2-4] In the past two decades, G4s have gained considerable attention due to their significant stability under physiological conditions. In addition, their formation in cells has been implicated in several key biological processes.[5,6] It has been suggested that G4 structures formed at telomeric ends of chromosomal DNA and stabilized by specific ligands might inhibit the activity of telomerase or lead to telomere dysfunction.[7-11] Telomerase, a ribonucleoprotein, is the enzyme responsible for telomere maintenance by elongating the G-rich 3'-overhang.[12] Recent studies have shown that telomerase can recognize and extend the unwound 3'-end of intermolecular parallel G4 structures and not the intramolecular antiparallel conformation.[13] The occurrence of DNA and RNA G4 structures in human cells has been visualized utilizing structure-specific antibodies[14,15] and smart fluorescent probes.[16] Interestingly, recent studies have shown that conserved Zika virus RNA sequences adopt stable, parallel-stranded G4 folds.[17] Therefore, G4-binding molecules may be utilized to inhibit viral replication.

Under acidic conditions, the complementary C-rich strand folds into i-motif structures consisting of two parallel duplexes intercalated in an antiparallel orientation and held together by hemiprotonated cytosine-cytosine ($C{\cdot}C^+$) base pairs (**Figure 4.1**).[18-20] Several factors have been found to stabilize these structures at physiological conditions such as molecular crowding and single-walled carbon nanotubes,[21] silver cations,[22] and negative supercoiling.[23] We have shown in Chapter 2 that 2'F-araC substitutions (**Figure 4.2**), significantly stabilize telomeric and centromeric i-motif structures at neutral pH.[24] To date, various i-motif structures had been reported in numerous oncogene promoter regions, such as c-MYC,[25] BCL-2,[26,27] VEGF,[28] and RET,[29] in addition to telomeric[30] and centromeric regions.[31,32] Furthermore, two recent reports show that several genomic sequences fold into i-motif structures at neutral pHs.[33,34] Even more recently, it has been demonstrated that i-motif structures in a DNA template can cause Klenow fragment DNA polymerase to stall immediately before the i-motif-forming region, suggesting that i-motifs could impede DNA replication or repair.[35]

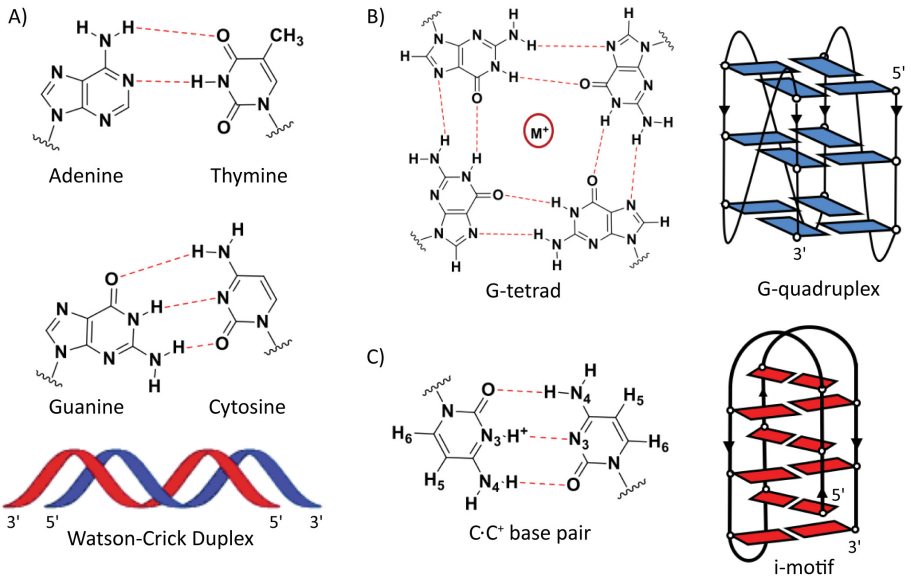


Figure 4.1: A) A·T and C·G Watson-Crick base pairs (top) and schematic representation of an antiparallel duplex (bottom). The G-rich strand is represented in blue and the C-rich strand in red. B) G-quartet stabilized by monovalent cation (left) and schematic representation of an intramolecular G-quadruplex (right). C) $C \cdot C^{+}$ base pair (left) and schematic representation of an intramolecular i-motif (right).

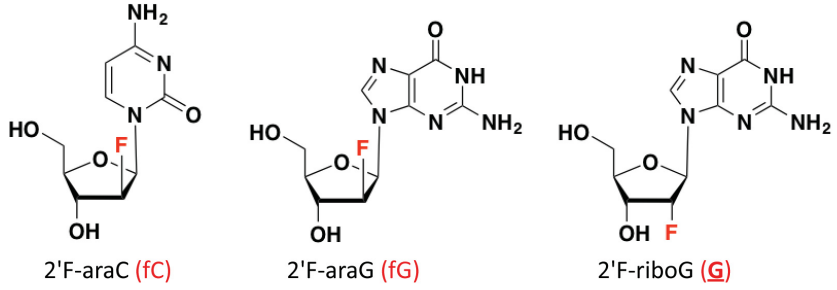


Figure 4.2: Structures of 2'-deoxy-2'-fluoro-arabinocytidine (2'F-araC, left), 2'-deoxy-2'-fluoro-arabinoguanosine (2'F-araG, middle) and 2'-deoxy-2'-fluoro-guanosine (2'F-riboG, right).

The telomeric G-rich 3'-overhang is a substrate of telomerase; however, no extension products are observed for the C-rich telomeric repeat due to the lack of complementarity between the C-rich strand and the RNA template of telomerase.[36] In blunt-ended telomeric duplexes where the two strands are fully hybridized, the G-rich strand is no longer a substrate of the enzyme.[37,38] Therefore, it is important to determine whether the transient formation of i-motif structures in the C-rich strand would free up the G-rich strand for elongation by, or inhibition of telomerase. For this reason, G4/i-motif-duplex interconversion is a process of significant

importance.[39-43] The kinetics of G4/i-motif unfolding and of duplex formation at pH 5.0 were first studied by Mergny and Phan in 2002.[44] Several other reports regarding G4/i-motif-duplex competition have been conducted, confirming that the competitive equilibria between G4, i-motif, and WC duplexes are primarily driven by the nature of the cation, the pH, and the sequence.[45-48]

In this work, we aim to stabilize parallel intramolecular G4 structures and tune the G4/i-motif-duplex interconversion rate at neutral pH by stabilizing i-motif and G4 structures simultaneously. For this purpose, C- and G-rich sequences modified with 2'F-araC, 2'F-araG, and 2'F-riboG residues were synthesized (**Table 4.1** and **Figure 4.2**). 2'-deoxy-2'-F-arabinonucleic acids (2'F-ANA) modifications can modulate G4 stability and topology by favoring the parallel propeller G4 topology over competing conformers, while 2'F-ribose modifications (2'F-RNA) disrupt the conformation of the furanose ring leading to less stable quadruplexes.[49-51] We show that 2'F-araG and 2'F-araC modified oligonucleotides provide very stable G4 and i-motif structures in potassium buffer at neutral pH, and that these structures pose a large kinetic barrier for duplex formation.

Table 4.1: Thermal stability of the telomeric 22G[a] and 22C[b]-rich sequences (pH 7.0).

Code	Sequence (5'-3')	T_m	ΔT_m
22G0	A GGG TTA GGG TTA GGG TTA GGG	65.6 ± 0.1	-
22G1	A (fG)GG TTA (fG)GG TTA (fG)(fG)G TTA (fG)GG	74.2 ± 0.9	+8.6
22G2	A (fG)GG TTA (fG)(fG)G TTA (fG)GG TTA (fG)GG	73.8 ± 0.2	+8.2
22G3	A (fG)GG TTA (fG)(fG)G TTA (fG)(fG)G TTA (fG)GG	80.2 ± 0.5	+14.6
22G4	A GGG TTA **G**GG TTA GGG TTA **G**GG	53.4 ± 0.2	-12.2
22G6	A GGG TTA **G**GG TTA GG**G** TTA **G**GG	57.6 ± 0.2	-8.0
22G7	A GGG TTA **GG**G TTA GG**G** TTA **G**GG	59.3 ± 0.2	-6.3
22C0	CCC TAA CCC TAA CCC TAA CCC T	n.d.	-
22C1	C(fC)(fC) TAA C(fC)(fC) TAA C(fC)(fC) TAA C(fC)(fC) T	25.7 ± 0.1	-
22C2	(fC)(fC)(fC) TAA (fC)(fC)(fC) TAA (fC)(fC)(fC) TAA (fC)(fC)(fC) T	30.0 ± 0.3	-

[a]The T_ms of the 22G-rich sequences were determined at 295 nm and 0.2 °C/min in 20 mM KP_i and 70 mM KCl buffer, pH 7.0 at 10 µM strand concentration. [b]The T_ms of the 22C-rich sequences were determined at 265 nm and 0.5 °C/min in 20 mM KP_i and 70 mM KCl buffer, pH 7.0 at 4 µM strand concentration. (fG): 2'F-araG, **G**: 2'F-riboG, (fC): 2'F-araC, and n.d. not detected.

4.2 Biophysical Properties of 22mer Cytosine and Guanine-Rich Single Strands and Duplexes

As discussed in the introduction of this thesis, G4, i-motif, and WC duplexes can be distinguished *via* nuclear magnetic resonance (NMR) since they possess distinct imino proton chemical shifts that originate from different base-pairing configurations.[52] Hence, the imino protons of base pairs stabilizing i-motifs ($C \cdot C^+$), duplexes (A·T and C·G), and G-tetrads appear at 15-16 ppm, 12-14 ppm, 10-12 ppm, respectively. The relative stability of these structures is readily assessed *via* UV-melting temperature experiments. Melting of duplex, i-motif, and G4 structures were monitored at 260 nm, 265 nm, and 295 nm, respectively.[53,54] Like NMR, circular dichroism (CD) provides a convenient means to distinguish between these structures since their spectra exhibit characteristic bands.[55-56] Therefore, NMR, UV-melting, and CD experiments served as complementary techniques to assess the stability and determine the kinetics of G4/i-motif-duplex interconversion (**Figure 4.3**). All the NMR spectra reported in this chapter were performed in collaboration with Prof. Carlos González at the CSIC in Madrid.

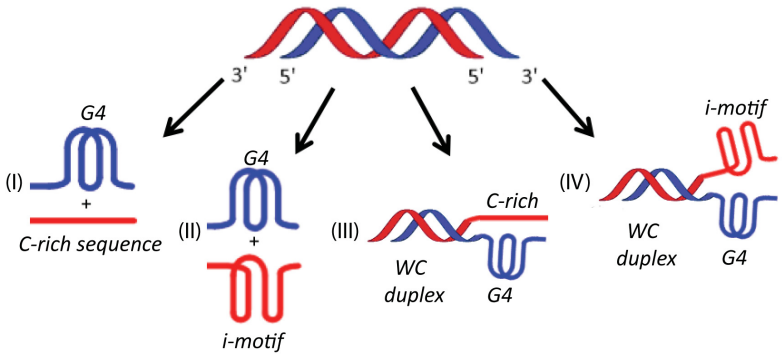


Figure 4.3: Models representing G4/i-motif-duplex interconversion of (I) 22mer human telomeric repeat forming G4, (II) 22mer human telomeric repeat forming G4 and i-motif secondary structures due to 2'F-ANA modifications, (III) 35mer human telomeric repeat forming a duplex and G4, and (IV) 35mer human telomeric repeat forming a duplex, a G4, and an i-motif structure within the same complex.

4.2.1 Structural Properties of 22mer Cytosine and Guanine-Rich Telomeric Single Strands

^{1}H-NMR and CD experiments performed on the single-stranded G-rich sequences confirm the formation of G4 structures. In all cases, buffers contained 20 mM KP_i and 70 mM

KCl at pH 7.0 (unless explicitly noted otherwise). ^{1}H-NMR spectra of 2'F-araG modified strands show signals in the 10 – 12 ppm range, which are characteristic of G4 formation (**Figure 4.4**). These imino signals are still observed at 45 °C indicating the formation of thermally stable G4 conformations. The CD spectrum of 22G0 shows two maxima indicating the presence of two conformers: an antiparallel conformation with a positive band at 295 nm and a parallel conformer with a positive band at 260 nm (**Figure 4.5A**). The incorporation of 2'F-araG and 2'F-riboG modifications triggers a shift in the spectra leading to a major positive signal at 260 nm and a negative signal at 240 nm, characteristic of a single parallel G-quadruplex (**Figure 4.5A**).

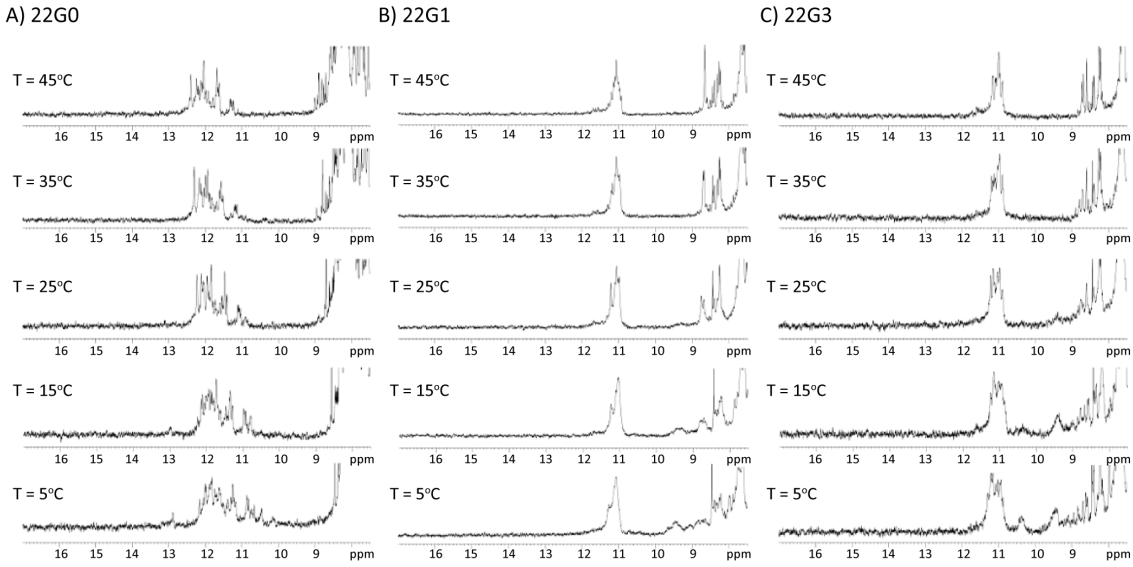


Figure 4.4: ^{1}H-NMR melting temperature experiments for A) 22G0, B) 22G1, and C) 22G3 in 20 mM KP_i and 70 mM KCl, pH 7.1.

Among the 2'F-araG modified sequences, the CD spectrum of 22G2 reveals a small shoulder at 295 nm suggesting the presence of a minor species folded in an antiparallel fashion. All 2'F-araG modified strands were found to be more stable compared to the control (22G0), as indicated by the increase in melting temperature, with the highest T_m observed for 22G3 (ΔT_m = +14.6 °C) (**Table 4.1** and **Figure 4.6**). 2'F-riboG modified G-rich strands are less stable compared to the unmodified sequence (**Table 4.1**).

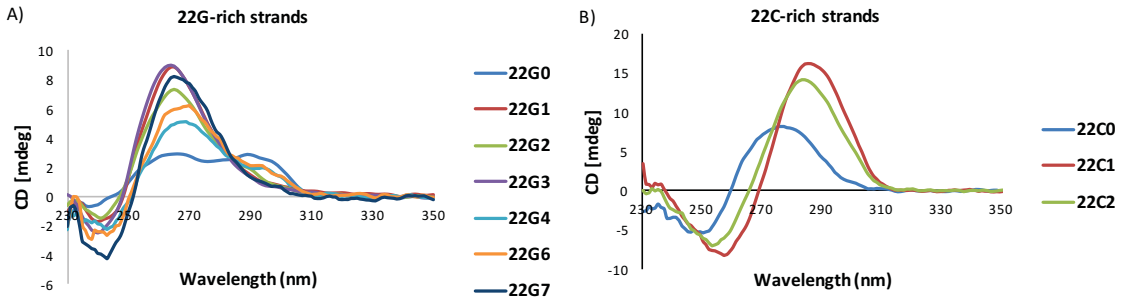


Figure 4.5: A) CD spectra for G-rich sequences recorded at 25 °C in 20 mM KP_i and 70 mM KCl buffer (pH 7.0). Oligonucleotide concentration is 10 µM in single strands. B) CD spectra for C-rich sequences recorded at 5 °C in 20 mM KP_i and 70 mM KCl buffer (pH 7.0). Oligonucleotide strand concentration is 20 µM in single strands.

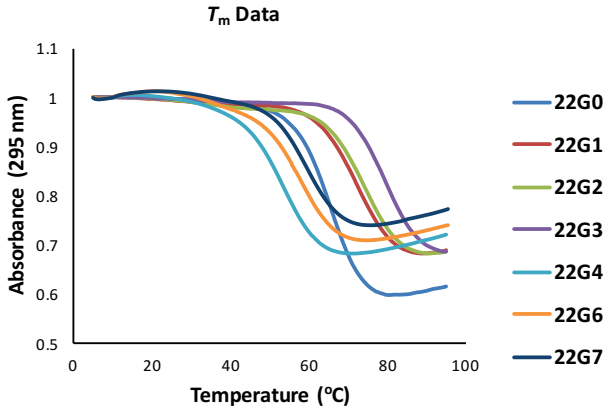


Figure 4.6: Thermal melting temperature experiments for the 22G-rich strands. Oligonucleotide concentration is 10 µM in single strands. T_m spectra were recorded at 295 nm in 20 mM KP_i and 70 mM KCl buffer, pH 7.0 at a heating rate of 0.2 °C/min.

UV-Melting experiments performed at different oligonucleotide concentrations (2 µM to 100 µM) had little effect on the T_m, thus confirming the formation of monomeric G4 structures in the case of 22G1 and 22G3 (**Table 4.2**). The presence of multiple conformations in the case of 22G2 and the decreased stability of the 2'F-riboG modified G-rich strands led us to exclude these sequences from further interconversion studies. At neutral conditions, the unmodified single-stranded C-rich sequence does not form stable i-motif structures even at high concentrations (100 µM, **Table 4.2**). However, the presence of 2'F-araC modifications provokes significant i-motif stabilization, in agreement with findings reported in Chapters 2 and 3.[24] The most stable

construction, 22C2 (T_m = 30.0 °C), was utilized in the G4/i-motif-duplex interconversion studies. 22C2 adopts an intramolecular folding topology since the T_m obtained from CD melting temperature experiments at 2 μM, 40 μM, and 100 μM was approximately 30.0 °C (**Table 4.2**).

Table 4.2: Melting temperatures (°C) of the telomeric 22G-rich and 22C-rich sequences (pH 7.0) at different concentrations.[a]

Code	Sequence (5'-3')	T_m 2 μM	T_m 10 μM	T_m 40 μM	T_m 50 μM	T_m 100 μM
22C0	CCC TAA CCC TAA C CCC TAA CCC T	n.d.	-	-	-	n.d.
22C2	((fC)(fC)(fC) TAA)$_3$ (fC)(fC)(fC) T	29.6	-	30[b]	-	30[b]
22G0	A GGG TTA GGG TTA GGG TTA GGG	66.5	65.6	66.5	66.5	66[c]
22G1	A (fG)GG TTA (fG)GG TTA (fG)(fG)G TTA (fG)GG	74.0	74.2	74.5	75.0	76[b]
22G3	A (fG)GG TTA (fG)(fG)G TTA (fG)(fG)G TTA (fG)GG	79.6	80.2	81.1 79[b]	81.6	81[b]
22G4	A GGG TTA **G**GG TTA GGG TTA **G**GG	-	53.4	54[b]	-	53[b]

[a]The T_ms of the 22G-rich sequences were recorded at 295 nm and 0.5 °C /min in 20 mM KPi and 70 mM KCl buffer (pH 7.0). The T_ms of the 22C-rich sequences were recorded at 265 nm and 0.5 °C /min in 20 mM KP_i and 70 mM KCl buffer (pH 7.0). (fG): 2'F-araG, **G**: 2'F-riboG, (fC): 2'F-araC, and n.d. not detected.
[b]CD melting temperature experiments in 20 mM KP_i and 70 mM KCl buffer (pH 7.0) at a ramp of 0.5 °C /min. T_ms were calculated from the plot of ellipticity at 265 nm as a function of temperature.
[c]T_ms were calculated from the plot of ellipticity at 295 nm as a function of temperature. A clear band was observed at 265 nm; however, the signal-to-noise ratio did not allow the extraction of a melting temperature at 265 nm.

4.2.2 Structural Properties of Cytosine and Guanine-Rich Telomeric Duplexes (Annealing *versus* Mixing)

UV-melting temperature (monitored at 260 nm) and circular dichroism experiments were performed at pH 5.8, 6.2, 6.6, 7.0 and 7.4 after annealing equimolar quantities of the 22C- and 22G-rich strands (*i.e.* a 1:1 mixture of the complementary strands was heated to 90 °C for 10 minutes and then slowly cooled down to room temperature, mix-heat-cool). Most of the melting curves display more than one transition (**Tables 4.3**). Looking closely at the melting curves (**Figure 4.7**) two hyperchromic and one hypochromic transitions are observed: a major, pH-independent melting transition corresponding to duplex unfolding and two minor transitions associated with the unfolding of i-motif (lowest pH-dependent T_m values) and G4 structures (highest pH-independent T_m values).

To confirm that the minor transitions correspond to the formation/melting of i-motif and G4 structures, UV-melting experiments for the constituent isolated 22G- and 22C-rich strands were performed at different pH values (**Table 4.4**). The T_m values observed for 22C0 and 22C2 decrease as the pH of the sample increases from 5.8 to 7.4 and were very similar to the values obtained for the first transition in the duplex melting experiments (**Tables 4.3** and **4.4**). Similarly, the T_m values observed for the G-rich sequences were comparable to those obtained for the hypochromic transition, confirming that this transition corresponds to the melting of G4 structures.

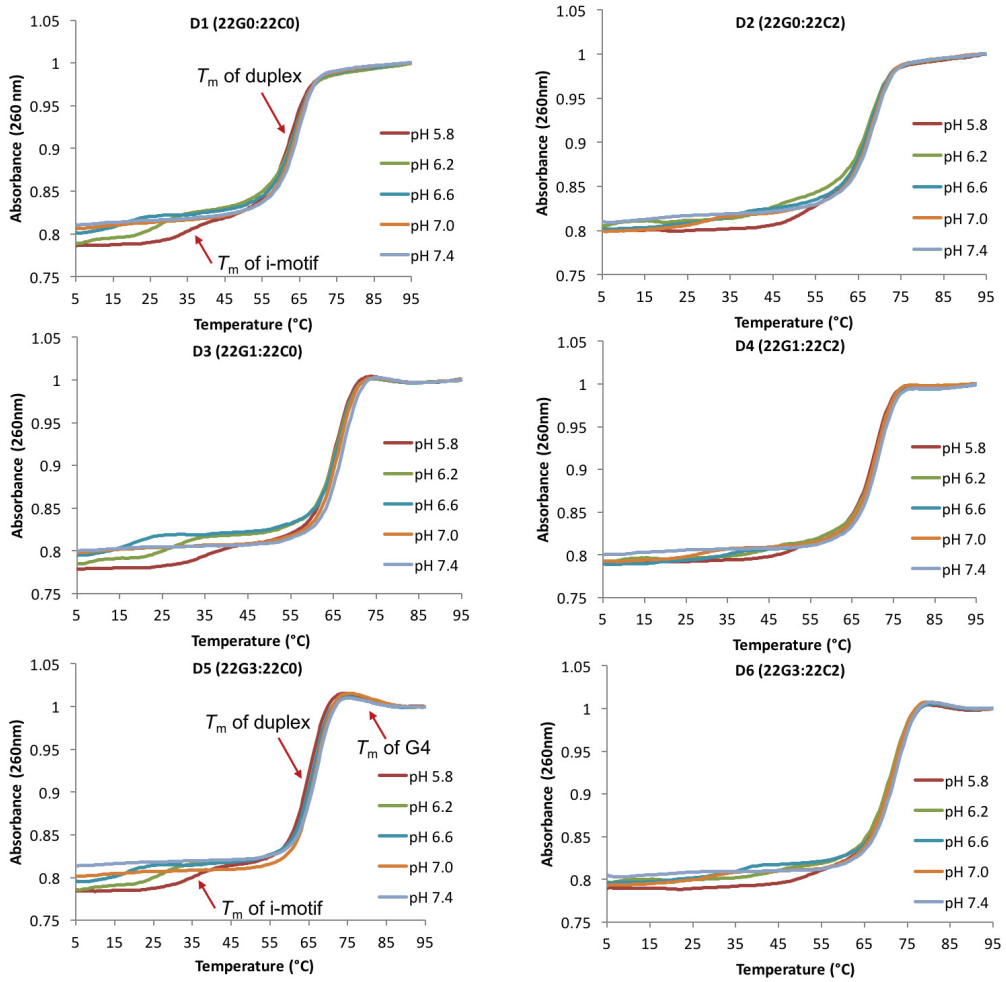


Figure 4.7: UV-melting curves of the 22mer duplexes. Duplex concentration is 2 μM in 20 mM KP_i and 70 mM KCl. Melting curves were recorded at a heating rate of 0.5 °C/min and at 260 nm. T_m values were calculated based on the first derivative method.

Table 4.3: Melting temperatures (°C) of 22mer human telomeric duplexes at different pHs.

Code	Duplex	T_m pH 5.8	T_m pH 6.2	T_m pH 6.6	T_m pH 7.0	T_m pH 7.4
D1	22G0:22C0	36.0 ± 0.8 63.3 ± 0.2	27.5 ± 0.3 64.2 ± 0.2	18.5 64.3 ± 0.2	10.8 ± 0.2 64.7 ± 0.2	11.0 65.5 ± 0.3
D2	22G0:22C2	49.5 68.5 ± 0.3	45.5 68.7 ± 0.3	35.0 ± 1.0 68.7 ± 0.3	26.5 ± 1.5 69.2 ± 0.3	19.5 69.5 ± 0.1
D3	22G1:22C0	35.4 ± 0.8 66.2 ± 0.2 78.5 ± 0.2	26.6 66.6 ± 0.2 78.5 ± 0.2	17.6 66.9 ± 0.3 78.4 ± 0.3	11.1 67.0 ± 0.3 78.1 ± 0.4	 67.7 ± 0.2 79.1 ± 0.1
D4	22G1:22C2	52.7 71.6 ± 0.5 83.5 ± 0.5	45.2 71.9 ± 0.4 83.3 ± 0.3	37.2 71.9 ± 0.4 83.3 ± 0.3	29.2 72.1 ± 0.4 83.5	 72.2 ± 0.3 84.0 ± 0.5
D5	22G3:22C0	36.0 ± 0.4 64.8 ± 0.1 80.1 ± 0.3	27.6 ± 0.6 65.8 ± 0.1 80.8 ± 0.2	17.3 ± 0.8 65.8 ± 0.1 80.3 ± 0.3	11.3 ± 0.6 66.3 ± 0.1 80.8 ± 0.1	 66.6 ± 0.2 80.8 ± 0.2
D6	22G3:22C2	 71.7 ± 0.2 84.4 ± 0.2	44.2 ± 0.4 71.7 ± 0.2 83.9 ± 0.5	34.9 ± 1.1 72.4 ± 0.4 84.4 ± 0.4	26.4 ± 1.3 71.9 ± 0.2 83.8 ± 0.3	21.0 72.6 ± 0.1 84.7 ± 0.1

Duplex concentration is 2 µM. T_m were recorded at 260 nm in 20 mM KP_i and 70 mM KCl. Melting curves were recorded at a heating rate of 0.5 °C/min.

Table 4.4: Melting temperatures (°C) of the telomeric 22G-rich and 22C-rich single strands at different pH values.

Code	Sequence (5'-3')	T_m pH 5.8	T_m pH 6.2	T_m pH 6.6	T_m pH 7.0	T_m pH 7.4
22G0	A GGG TTA GGG TTA GGG TTA GGG	71.2 ± 0.2	71.0 ± 0.6	70.7 ± 0.4	70.8 ± 0.3	71.0 ± 0.3
22G1	A (fG)GG TTA (fG)GG TTA $(fG)_2$G TTA (fG)GG	74.1 ± 0.5	73.6 ± 0.8	73.2 ± 0.9	74.0 ± 0.3	72.7 ± 0.9
22G3	A (fG)GG TTA $(fG)_2$G TTA $(fG)_2$G TTA (fG)GG	79.3 ± 0.9	79.5 ± 0.7	78.8 ± 0.9	79.6 ± 0.5	79.5 ± 0.3
22G4	A GGG TTA **G**GG TTA GGG TTA **G**GG	53.8 ± 0.3	53.0 ± 0.1	53.8 ± 0.3	53.3 ± 0.3	53.3 ± 0.3
22C0	CCC TAA CCC TAA CCC TAA CCC T	36.0 ± 0.1	28.1 ± 0.1	18.7 ± 0.2	n.d.	n.d.
22C2	$(fC)_3$ TAA $(fC)_3$ TAA $(fC)_3$ TAA $(fC)_3$ T	51.8 ± 0.2	44.5 ± 0.2	34.8 ± 0.2	29.6 ± 0.1	21.4 ± 0.3

Oligonucleotide concentration is 2 µM in 20 mM KP_i and 70 mM KCl. Melting curves were recorded at a heating rate of 0.5 °C/min and at 260 nm. (fG): 2'F-araG, **G**: 2'F-riboG, (fC): 2'F-araC, and n.d. not detected.

CD experiments as a function of pH were performed for all the samples containing 2'F-ANA modified C- and G-rich sequences (**Figure 4.8**). The samples containing 22C0 hybridized to 22G0 or to 2'F-ANA modified G4s, all exhibit a clear positive band at 285 nm at pH 5.8 and 6.2. This band decreases in magnitude as the pH increases to 7.0 and 7.4, reflecting the progressive unfolding of the i-motif structure with increasing pH. However, in the case of 22C2, this band was found in the CD spectra up to pH 7.0. Moreover, in all of the studied samples, a positive band appears between 270-260 nm with a negative band around 240 nm. To

unequivocally determine which species were forming under these conditions, 1D NMR spectra were recorded at neutral pH (**Figure 4.9**). The spectra obtained after mixing and annealing equimolar quantities of the 22G-rich and 22C-rich strands presented imino signals in the 12-14 ppm region, indicating that the formation of WC duplex structures is favored under slow annealing conditions and at higher concentrations, which is expected since duplex formation is concentration-dependent (**Figure 4.9**). The well-known ability of 2'F-ANA modifications to stabilize duplex structures can be noticed in the presence of clear imino signals for the modified duplexes at higher temperatures.[57] This is also consistent with the higher T_m values of the main melting transition in the modified duplexes (**Table 4.3** and **Figure 4.9**).

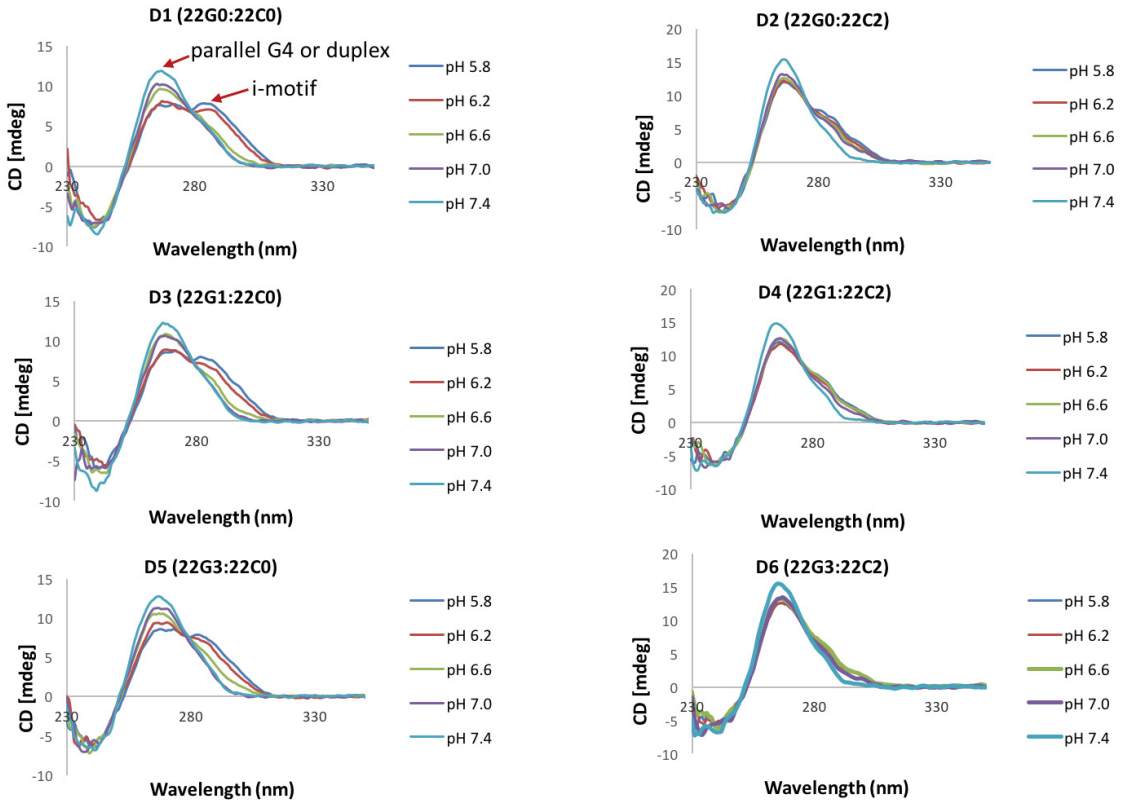


Figure 4.8: CD spectra as a function of pH for all the samples containing 2'F-ANA modified C- and G-rich sequences. Oligonucleotide concentration is 15 µM. CD spectra recorded at 5 °C in 20 mM KP_i and 70 mM KCl.

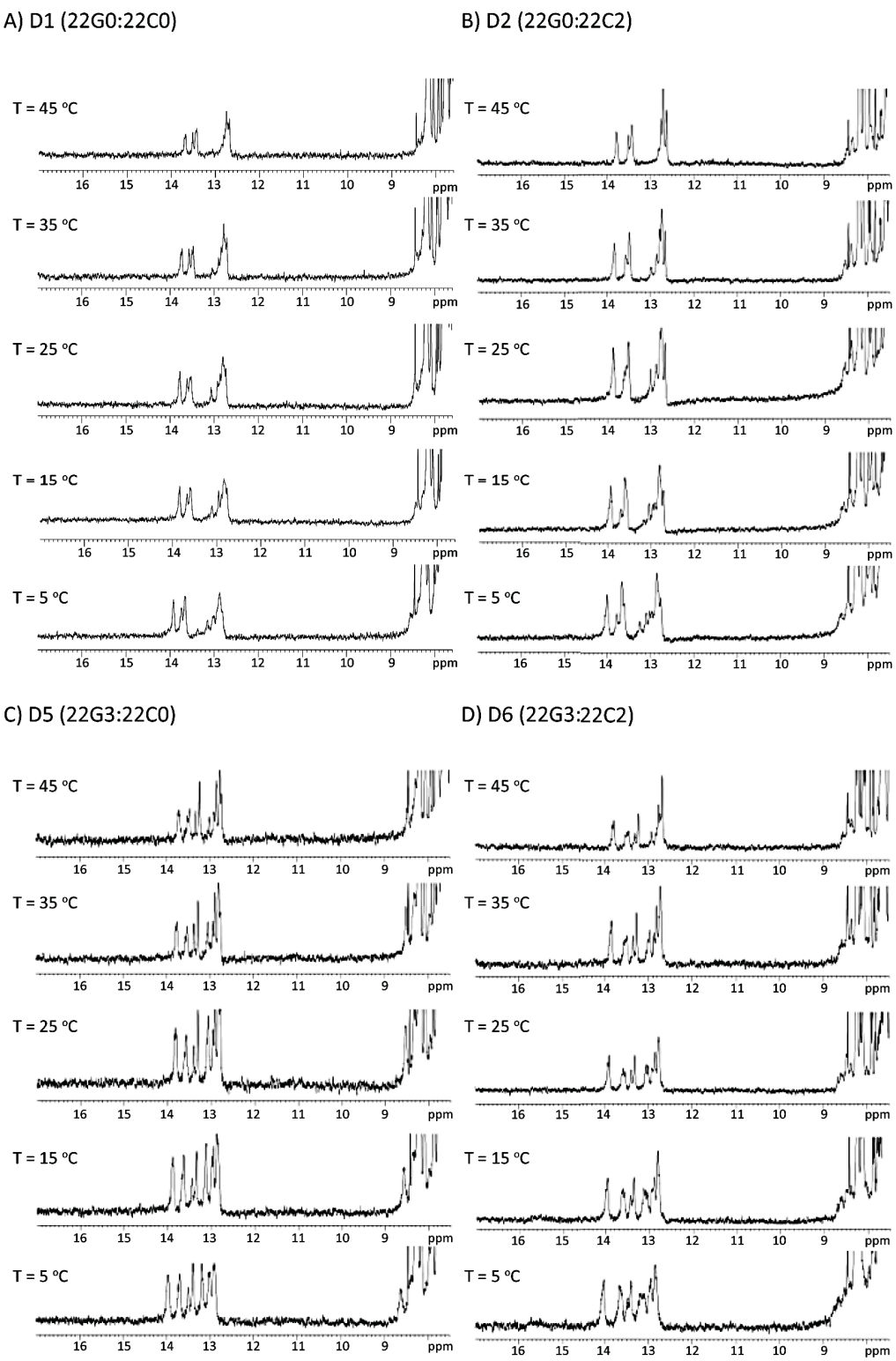


Figure 4.9: ^{1}H-NMR melting temperature experiments at pH 7.0 of A) D1 (22G0:22C0), B) D2 (22G0:22C2), C) D5 (22G3:22C0), and D) D6 (22G3:22C2). Samples were annealed for 10 minutes at 90 °C and stored at 4 °C before the NMR melting temperature experiments were performed.

4.2.3 NMR Kinetic Experiments of 22mer Mixes

The kinetics of duplex formation were monitored by following the signal changes in the imino region of the NMR spectra (**Figures 4.10** and **4.11**). The 22G- and 22C-rich single strands were annealed separately and then mixed in equimolar amounts without further annealing (heat-cool-mix). Changes in the relative populations of duplex, G4, and i-motif species over time were followed by observing the change in their relative signal intensities. In the case of the control 22mer sample (D1), the intensity of G4 imino signals decreases within hours at 5 °C and within minutes at 25 °C. After 20 hours (5 °C, **Figure 4.10**) and 2 hours (25 °C, **Figure 4.11**), the G4 imino signals vanish completely, indicating that the two strands are fully hybridized.

Incorporation of 2'F-ANA modifications in any of the strands slows down the hybridization process dramatically. Indeed, at 5 °C, when only the C-strand is modified (D2, **Figure 4.10B**), i-motif signals are present after 5 days. Likewise, when only the G-strand is modified (D5, **Figure 4.10C**), G4 imino signals are visible after 35 days at 5 °C. Fluorination in the G-strand gives rise to G4 structures that are more stable than the duplexes. This is consistent with the T_m data (**Table 4.3**) where the G4 transition always occurred at higher temperatures relative to the duplex. This effect is even more pronounced when the two strands are modified (D6, **Figure 4.10D**); G4 and i-motif imino signals are observed after 35 days with only very minor WC imino signals detectable.

A comparison of the different kinetics of duplex formation at 25 °C indicates that full duplex formation is achieved after approximately 11 days when the C-strand alone is modified (D2, **Figure 4.11B**). When both strands are modified, the speed of duplex formation is inversely proportional to the number of 2'F-araG residues incorporated (**Figure 4.11C** *vs* **4.11D**). The G4 signals in D2 at 25 °C are sharp and co-exist with the WC imino signals for 11 days. On the other hand, the G4 signals of D2 at 5 °C are very broad and disappear after almost 20 hours. This might be due to aggregation of G4 at lower temperatures leading to broad imino signals of the unmodified G-rich strand. These broad signals tend to disappear at a faster rate compared to the G4 signals at 25 °C.

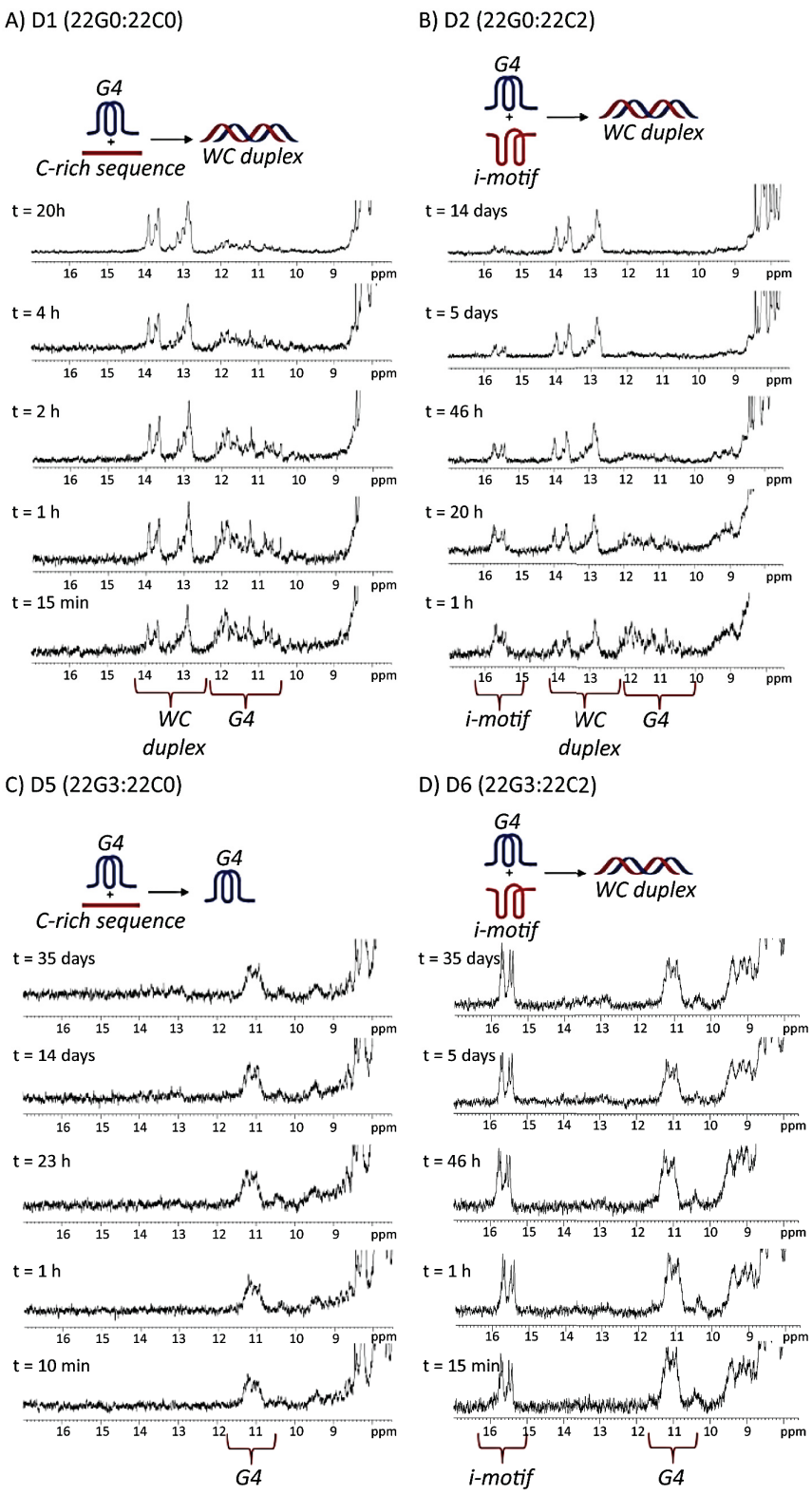


Figure 4.10: ^{1}H-NMR monitored kinetic experiments for A) D1 (22G0:22C0), B) D2(22G0:22C2), C) D5 (22G3:22C0), and D) (22G3:22C2) in 1:1 mixture of 22G:22C at pH 7.1 in 20 mM KP_i and 70 mM KCl at T = 5 °C.

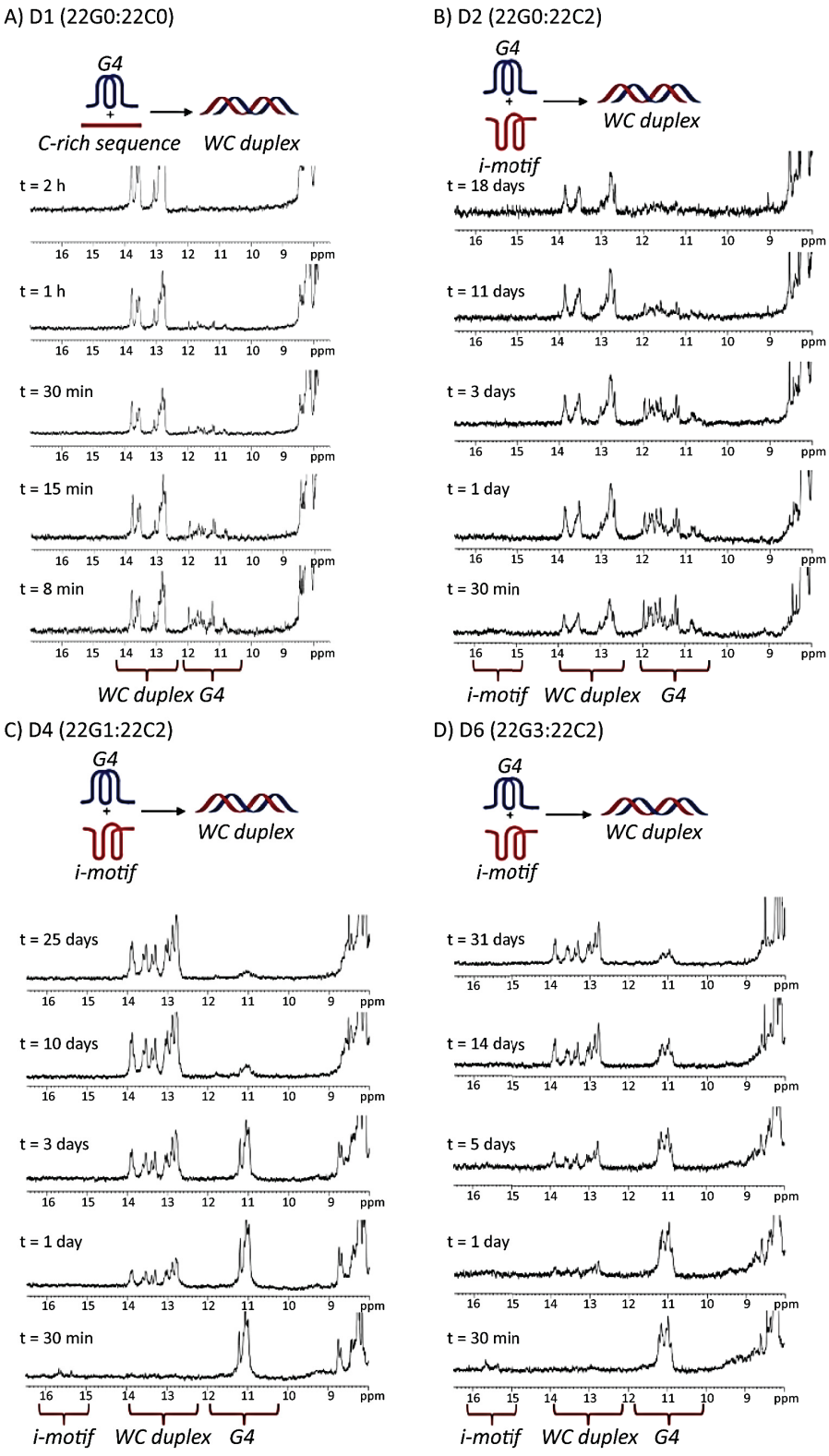


Figure 4.11: ^{1}H-NMR monitored kinetic experiments for A) D1 (22G0:22C0), B) D2(22G0:22C2), C) D4 (22G1:22C2), and D) (22G3:22C2) in 1:1 mixture of 22G:22C at pH 7.1 in 20 mM KP_i and 70 mM KCl at T = 25 °C.

When the G-strand is unmodified and the C-strand is modified (D2; 22G0:22C2), the rate-determining step for duplex formation at 5 °C is the unfolding of 2'F-ANA modified i-motif. When the G-strand is modified and the C-strand is unmodified (D5; 22G3:22C0), the unfolding of 22G3 is rate-determining. When both strands are modified (D6; 22G3:22C2), the i-motif and G4 structures persist with no duplex formation observed even after 35 days at 5 °C. Therefore, at 5 °C, G4 and/or i-motif unfolding are the rate-determining steps for duplex formation and the slowdown effect is exacerbated when modifications are incorporated in the G-strand rather than in the C-strand. When the temperature is raised to 25 °C, 22C2 is partially unfolded (T_m = 30 °C) and the G4/i-motif to duplex interconversion rate is dominated by G4 unfolding. Other studies have shown that i-motif unfolding represents a large kinetic barrier to duplex formation;[58] however, this is the first time that this effect has been demonstrated at neutral pH.

4.3 Telomerase Activity Assays

Telomerase activity assays were performed for the modified G4 structures and 22mer mixes in collaboration with Aaron Moye in Dr. Tracy Bryan's lab at the Children's Medical Research Institute in Sydney. The 18mer linear control refers to a telomeric oligonucleotide that has been denatured immediately prior to the telomerase activity assays and represents the extension of a linear substrate (lane 1, **Figure 4.12**). Pre-folded 22G0 exhibited the "stuttering" pattern revealing the typical poor extension observed for antiparallel G4 structures (lane 3, **Figure 4.12**).[13,59,60] However, interestingly, pre-folded 2'F-araG modified G4s formed from 22G1, 22G2, and 22G3 (lanes 4-6) were extended by telomerase with the normal 6-nt repeat pattern. Therefore, telomerase is able to extend intramolecular parallel G4s and thus our data further confirms that telomerase activity is highly dependent on the folding topology of G4 structures. In the case of the dimeric intermolecular parallel G4s, the Bryan group hypothesized that telomerase partially unwinds the G4 structure prior to extension. Therefore, investigating the mechanism by which telomerase elongates the highly stable modified G4s is of utmost importance.

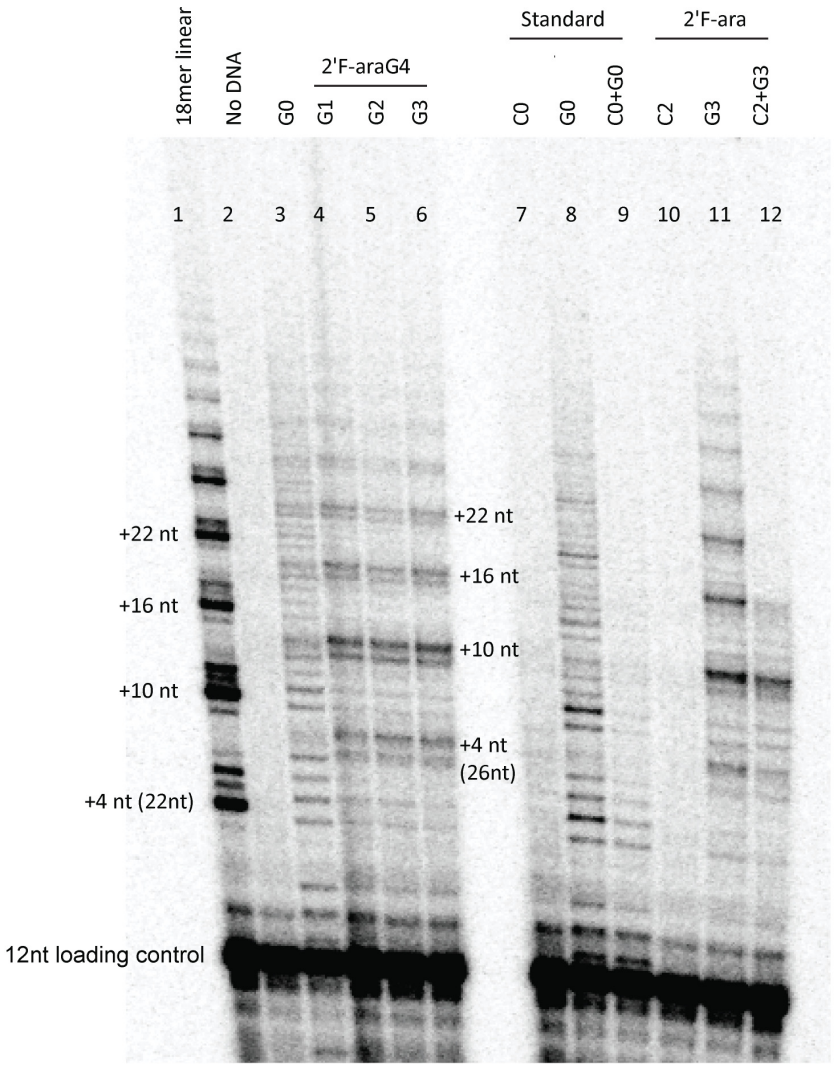


Figure 4.12: Telomerase activity assay with immunopurified telomerase at pH 7.5 and 37 °C. Loading control indicates a ^{32}P-labelled oligonucleotide used as a control for loading and recovery. Samples were pre-folded or annealed at pH 7.0 in 20 mM KP_i + 20 mM KCl. Lane 1: denatured 18mer linear control, lane 2: no DNA added, lane 3: pre-folded 22G0, lane 4: pre-folded 22G1, lane 5: pre-folded 22G2, lane 6: pre-folded 22G3, lane 7: pre-folded 22C0, lane 8: pre-folded 22G0, lane 9: mixing pre-folded 22G0 + pre-folded 22C0, lane 10: pre-folded 22C2, lane 11: pre-folded 22G3, lane 12: mixing pre-folded 22G3 + pre-folded 22C2.

For the C-rich strands, no extension of 22C0 (C-rich negative control, lane 7) and 22C2 (2'F-ANA modified C-rich, lane 10) strands was observed, as expected.[36] As observed in **Figure 4.12**, mixing of pre-folded 22G0 + pre-folded 22C0 showed very little activity either due to duplex formation, which is not expected to be substrate of telomerase, or the single-stranded C-rich sequence inhibited telomerase activity.[37,38] Interestingly, the addition of pre-folded 22C2 to pre-folded 22G3 inhibited the extension of 22G3 by telomerase (lane 12 *versus* lane 11, **Figure 4.12**). A + 10-nt band was observed in lane 12, indicating only partial extension of the stable G4 species. Since the T_m of 22C2 at pH 7.0 is 30.0 °C therefore, at 37 °C, temperature at which the telomerase activity assay was initially performed, the i-motif structure is partially unfolded.

We hypothesize that the C-rich DNA might be directly inhibiting telomerase since, at 37 °C, 22G3 is expected to maintain a quadruplex topology since it exhibits a T_m of 80.2 °C. Therefore, a linear single-stranded G-rich sequence will not be available to hybridize to the unfolded C-rich complementary strand for duplex formation. Further experiments need to be performed in order to determine whether the single-stranded C-rich portion or the folded i-motif structure causes telomerase inhibition.

The telomerase activity assays were repeated at 25 °C, to minimize 22C2 unfolding during the telomerase activity assay (**Figure 4.13**). Again we observed that the mixture of pre-folded 22C2 + 22G3 (*i.e.*, G4 + i-motif, lane 8) was also poorly extended by telomerase, similar to the results obtained at 37 °C. This implies that (22C2 + 22G3) is either forming a duplex, which is unlikely under these conditions, or the presence of 22C2 folded into an i-motif structure inhibited telomerase activity. These data are preliminary and the mechanism of inhibition is still under investigation to conclude whether i-motif formation in 22C2 has an impact on telomerase activity. Follow-up experiments that will be performed to test this possibility are described in Chapter 5.

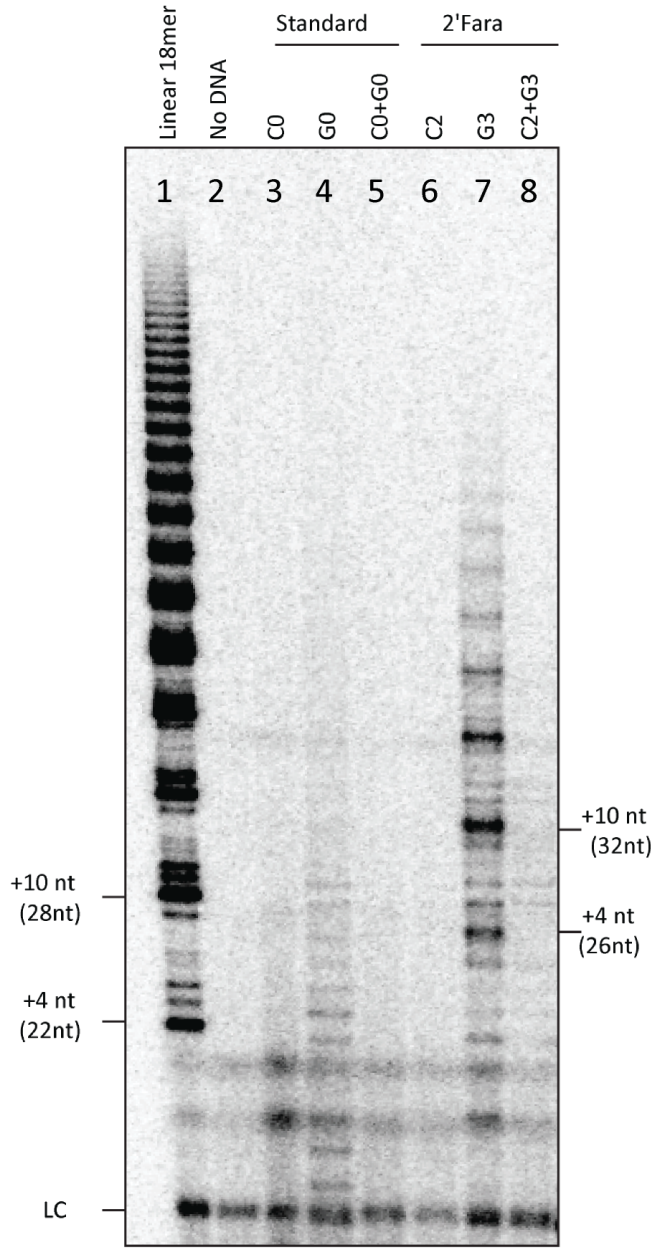


Figure 4.13: Telomerase activity assay with immunopurified telomerase at pH 7.5 and 25 °C. Loading control indicates a ^{32}P-labelled oligonucleotide used as a control for loading and recovery. Samples were pre-folded or annealed at pH 7.0 in 20 mM KP_i + 20 mM KCl. Lane 1: denatured 18mer linear control, lane 2: no DNA added, lane 3: pre-folded 22C0, lane 4: pre-folded 22G0, lane 5: mixing pre-folded 22G0 + pre-folded 22C0, lane 6: pre-folded 22C2, lane 7: pre-folded 22G3, lane 8: mixing pre-folded 22G3 + pre-folded 22C2.

4.4 Properties of 35mer Cytosine and Guanine-Rich Single Strands and Duplexes

As mentioned in the introduction of the thesis, several recent studies suggest that G4 and i-motif structures are mutually exclusive.[40,41] That is, either a G4 or an i-motif structure can form at a certain instance in the same region on complementary strands of a duplex. This mutual exclusivity was hypothesized to be a result of steric hindrance that one tetraplex imposes to impede the formation of the other tetraplex on the complementary strand. Interestingly, G4 and i-motifs were formed simultaneously when arranged in a slightly offset fashion along the complementary stands of a DNA duplex,[39] confirming that steric hindrance might be the reason behind the mutual exclusivity in promoter regions. Therefore, in order to study a scenario where i-motif and G4 structures are in close proximity, longer constructs were prepared. These constructs consist of the same 22-nt sequences studied above, but linked to a 13-nt stretch of mixed base composition (**Table 4.5**, **Figure 4.3-III** and **4.3-IV**). A TTT spacer between the G4/i-motif domain and the duplex domain was inserted to allow for some structural flexibility and to prevent duplex destabilization caused by the proximal G4 and i-motif structures.[61]

Table 4.5: Melting temperatures (°C) of the telomeric 35G-rich and 35C-rich sequences.

Code	Sequence (5'-3')	T_m pH 7.0	ΔT_m pH 7.0	T_m pH 5.8	ΔT_m pH 5.8
35G0	C ACA GAT GCG TTT A GGG TTA GGG TTA GGG TTA GGG	56.0 ± 0.2	-	57.5 ± 0.1	-
35G3	C ACA GAT GCG TTT A (fG)GG TTA (fG)(fG)G TTA (fG)(fG)G TTA (fG)GG	75.0 ± 0.1	+19.0	75.7 ± 0.2	+18.2
35C0	CCC TAA CCC TAA CCC TAA CCC T TTT CGC ATC TGT G	n.d.	-	35.5 ± 0.1	-
35C2	$(fC)_3$ TAA $(fC)_3$ TAA $(fC)_3$ TAA$(fC)_3$ T TTT CGC ATC TGT G	31.3 ± 0.2	-	52.2 ± 0.1	+16.7

Oligonucleotide concentration is 4 μM in 20 mM KP_i and 70 mM KCl. Melting curves were recorded at a heating rate of 0.5 °C/min. T_m were recorded at 260 nm for C-rich sequences and 295 nm for G-rich sequences. (fG): 2'F-araG, (fC): 2'F-araC, and n.d. not detected.

Similar to the data obtained for the 22-nt single-stranded sequences, the CD spectra of the control G-rich strand (35G0) exhibits the characteristic signature of an antiparallel G4 structure, while the 2'F-araG modified strand (35G3) possesses a positive band at 260 nm and a negative band at 240 nm, indicating the formation of a parallel G4 (**Figure 4.14**). As expected, the

stability of the 35C-rich sequences is pH-dependent (**Table 4.5**). The CD spectra of the 2'F-araC modified structures at pH 7.0 indicate the formation of i-motifs.

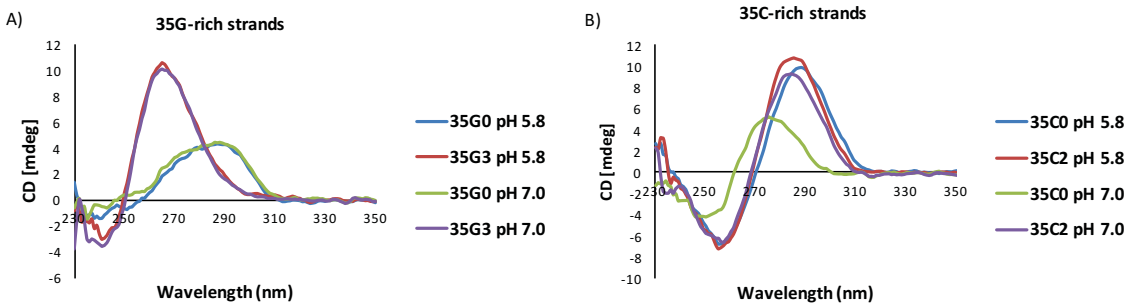


Figure 4.14: CD spectra for A) 35G-rich sequences at 25 °C and B) 35C-rich sequences at 5 °C. Oligonucleotide concentration is 30 µM in 20 mM KP_i and 70 mM KCl.

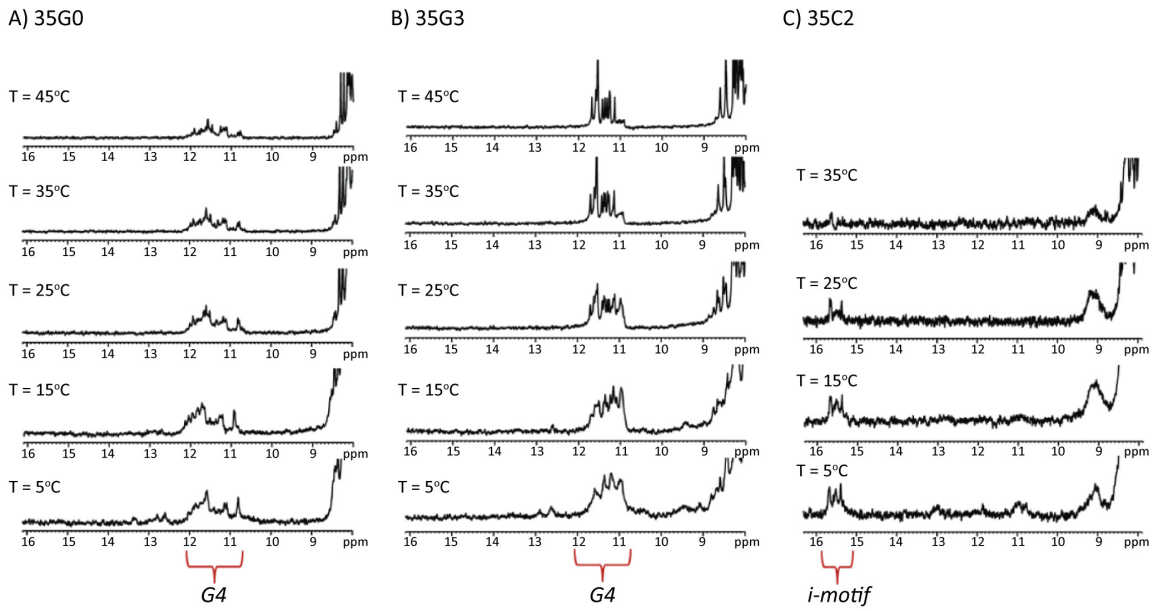


Figure 4.15: ^{1}H-NMR melting experiments at pH 7.1 in 20 mM KP_i and 70 mM KCl. G4 imino signals show at 10 – 12 ppm and i-motif imino signals at 15 – 16 ppm.

To confirm the formation of G4 and i-motif secondary structures in those constructs, ^{1}H-NMR melting experiments were performed (**Figure 4.15**) and revealed broad signals for the protons of 35G0, indicating a dynamic equilibrium between multiple conformers. On the other hand, 35G3 exhibits sharp and well-dispersed imino signals, in line with the formation of a single, parallel conformer as indicated by the CD spectra. In the case of 35C2, i-motif imino

signals are clearly observed up to 25-35 °C and at neutral pH. Therefore, the presence of single-stranded, 13-nt long overhangs does not prevent the formation of stable i-motif and G4 structures. When 35G- and 35C-rich strands were annealed, they hybridized to form WC duplexes. The UV melting curves show only one transition corresponding to duplex dissociation (**Figure 4.16**). The T_m values obtained are comparable with those obtained for the 22-nt duplexes (**Table 4.6**). This is likely due to the presence of three T·T mismatches within the duplexes which can be detected in the ^{1}H-NMR spectra as imino signals in the 10 - 11 ppm range (**Figure 4.17**).

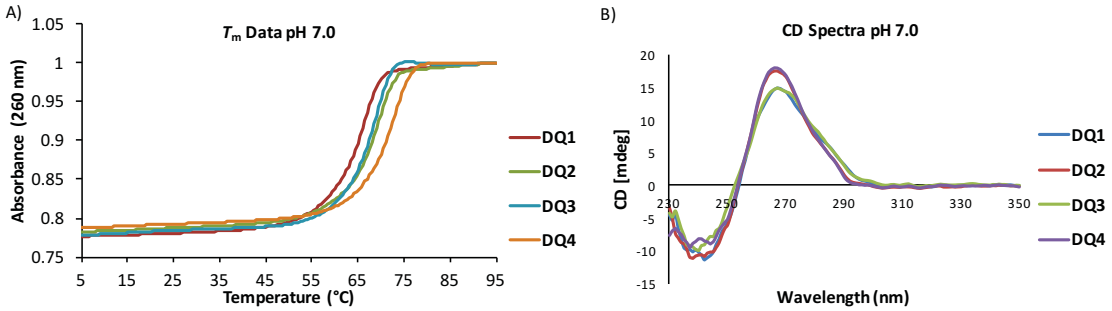


Figure 4.16: A) UV-melting curves of 35mer G- and C-rich annealed duplexes recorded at 260 nm and 0.5 °C/min in 20 mM KP_i, 70 mM KCl, pH 7.0 buffer. Oligonucleotide concentration is 2 µM. B) CD spectra of the same duplexes recorded at 5 °C in 20 mM KP_i and 70 mM KCl, pH 7.0 buffer. Oligonucleotide concentration is 15 µM.

Table 4.6: Melting temperatures (°C) of 35mer human telomeric duplexes at different pHs.

Code	T_m (pH 7.0)	ΔT_m (pH 7.0)	T_m (pH 5.8)	ΔT_m (pH 5.8)
DQ1 (35G0:35C0)	66.2 ± 0.2	-	65.5 ± 0.1	-
DQ2 (35G0:35C2)	69.5 ± 0.1	+3.3	69.4 ± 0.2	+3.9
DQ3 (35G3:35C0)	69.1 ± 0.3	+2.9	67.8 ± 0.2	+2.3
DQ4 (35G3:35C2)	72.9 ± 0.2	+6.7	72.5 ± 0.1	+7.0

Duplex concentration is 2 µM. Melting curves were recorded at 260 nm in 20 mM KP_i and 70 mM KCl and at a heating rate of 0.5 °C/min.

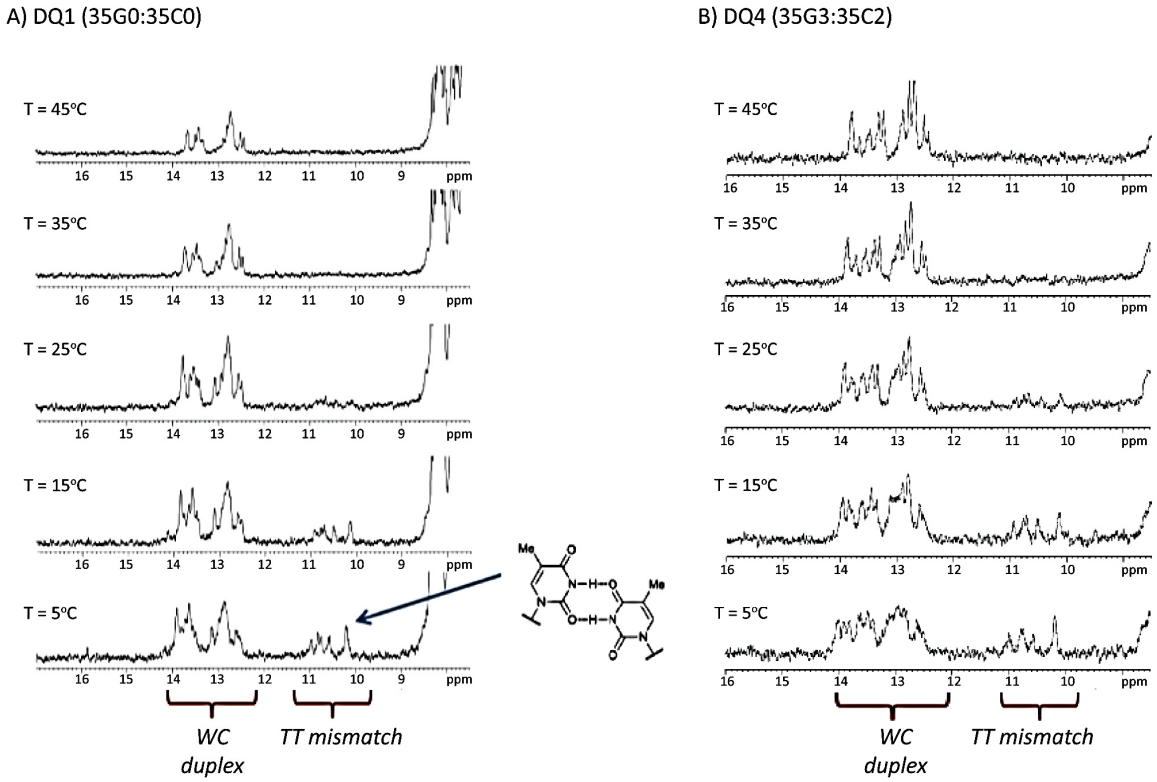


Figure 4.17: ^{1}H-NMR melting temperature experiments of A) 35G0:35C0 and B) 35G3:35C2 samples. Samples were annealed for 10 minutes at 90 °C and stored at 4 °C before the NMR melting temperature experiments were performed.

4.4.1 NMR Kinetic Experiments of 35mer Mixes

The kinetics of duplex formation upon mixing the pre-folded 35mer G4 and i-motif structures were monitored by ^{1}H-NMR spectroscopy (**Figure 4.18**). Some imino signals from WC base pairs are observed immediately after mixing, corresponding to hybridization of the mixed-base complementary ends. However, like the 22mer sequences, NMR experiments show that full hybridization of the 35 nucleotides takes much longer when 2'F-ANA nucleotides are included in any of the strands. For example, in the native 35mer duplex (35G0:35C0), full duplex formation is achieved after two days at T = 5 °C (**Figure 4.18A**). In DQ3, where the C-rich strand does not form stable i-motif structures and the G4 structure is very stable (T_m of 35G3 = 75.0 °C, **Table 4.5**), G4 and WC imino signals can still be observed after 26 days (**Figure 4.18C**). This is in agreement with the results obtained previously by Ren and co-authors on the same 22G-rich sequence attached to a Dickerson-Drew dodecamer tail, reporting "little or no cross talk" between the G4 structure and the proximal 12-nt duplex.[62]

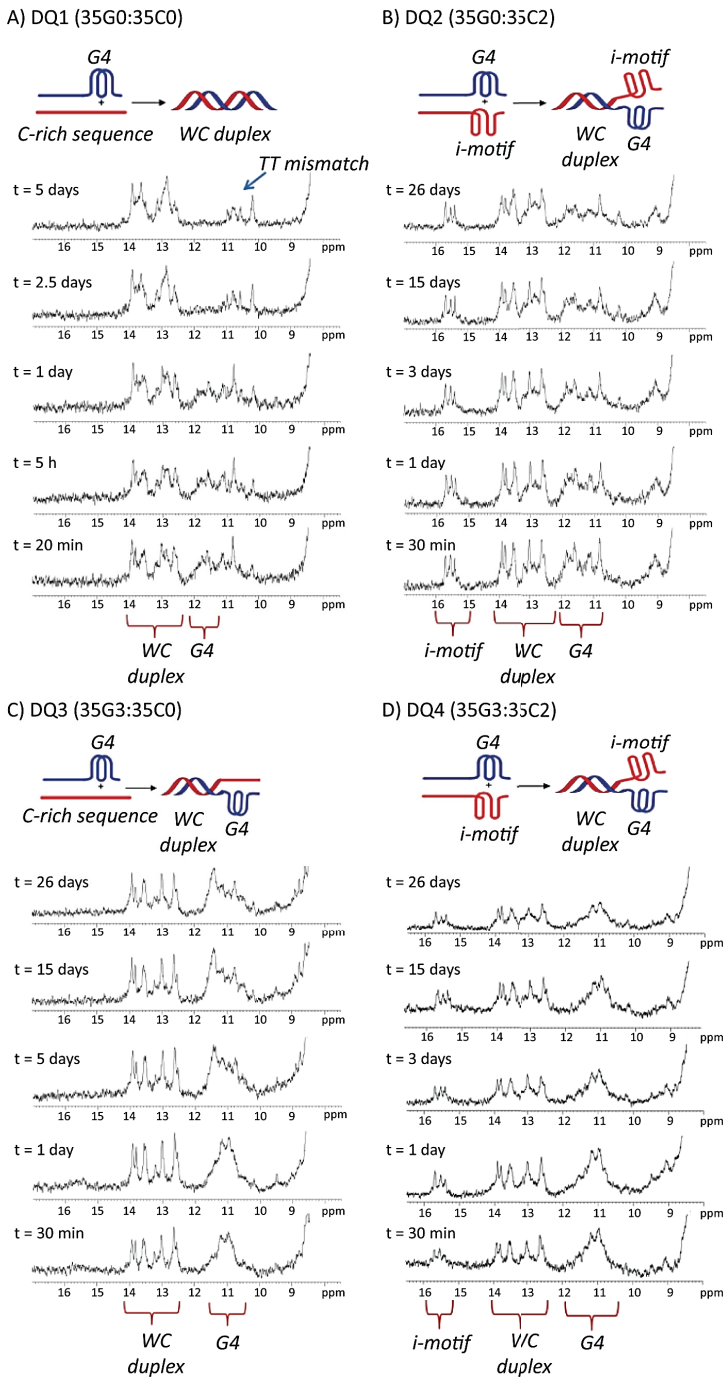


Figure 4.18: ^{1}H-NMR spectra acquired at different times and at 5 °C for a 1:1 mixture of A) DQ1 (35G0:35C0), B) DQ2 (35G0:35C2), C) DQ3 (35G3:35C0), and D) DQ4 (35G3:35C2). Buffer conditions: 20 mM KP_i and 70 mM KCl, pH = 7.1.

Therefore, the presence of G4 structures does not prevent duplex formation in the adjacent overhang. In cases where the C-rich strands form stable i-motif structures (**Figure 4.18B** and Figure **4.18D**), G4 and i-motif imino signals can be detected along with WC imino

signals arising from overhang hybridization. Very clear i-motif, G4, and WC imino signals can still be observed one month after mixing the pre-folded 35C-rich and 35G-rich strands. Therefore, given the long-term kinetic stability observed for the modified G4 and i-motif structures, 2'F-ANA modifications can be utilized as a convenient means to "trap" these structures at neutral conditions.

Next we wanted to confirm that the 1D NMR spectra observed in **Figure 4.18D** corresponds to a single hybrid complex wherein a DNA duplex, a G4, and an i-motif co-exist in the same complex at neutral pH (**Figure 4.19A**) rather than a mixture of complexes in which either the i-motif (B1, **Figure 4.19**) or the G4 (B2, **Figure 4.19)** are unfolded.

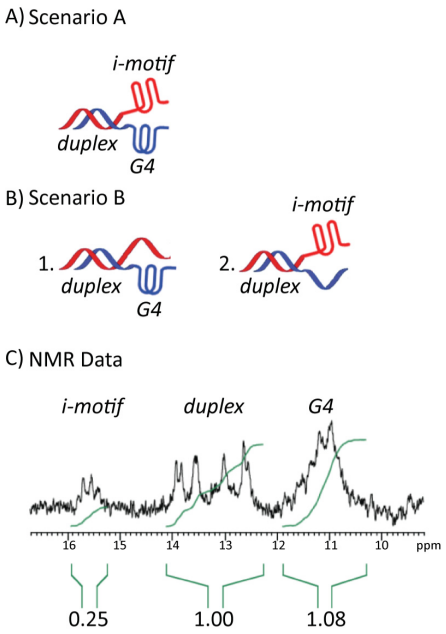


Figure 4.19: Schematic representation of the proposed scenarios: Scenario A) a duplex moiety co-exists in the same complex with G4 and i-motif. Scenario B) a duplex moiety coexists with a C-rich ssDNA and a G4 (B1), a G-rich ssDNA and an i-motif (B2). C) Imino region of the ^{1}H-NMR spectrum of 35G3:35C2 recorded at T = 5 °C after 30 min from mixing the two strands and showing the values resulting from the integration of i-motif, duplex, and G4 imino signals.

Gel electrophoretic experiments showed that all strands form dimeric complexes, without substantial presence of monomers or higher order structures (**Figure 4.20**). The pre-folded-mixed samples (lanes 10 - 13) show one band with very similar mobility to the annealed

duplexes. We do not expect full hybridization since under the gel conditions (pH 7.0, T = 10 °C), 35G0 and 35G3 exhibit high melting temperatures (56 °C and 75 °C, respectively) and therefore should be in the G4 folded state. Next, the 2D NMR data obtained indicate that the complex in which the three secondary structures occur simultaneously is significantly populated.

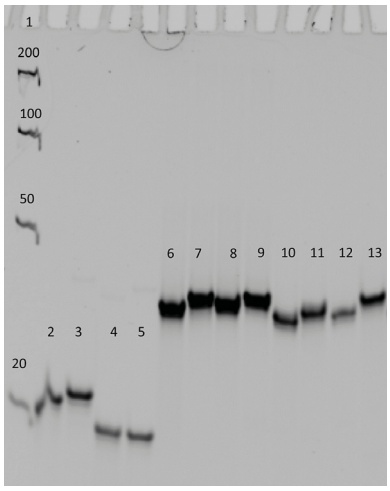


Figure 4.20: Gel electrophoresis experiments: 24 % polyacrylamide in TAE buffer pH 7.0, 10 °C, 4h 30 min, 280 V, and stained using SYBR Gold. Samples annealed in 20 mM KP_i + 70 mM KCl, pH 7.0. Lane 1: DNA ladder, lane 2: pre-folded 35G0, lane 3: pre-folded 35G3, lane 4: pre-folded 35C0, lane 5: pre-folded 35C2, lane 6: annealed 35G0:35C0, lane 7: annealed 35G0:35C2, lane 8: annealed 35G3:35C0, lane 9: annealed 35G3:35C2. Samples pre-folded then "mixed" for 18 hours before running the gel: lane 10: 35G0:35C0, lane 11: 35G0:35C2, lane 12: 35G3:35C0, lane 13: 35G3:35C2.

Considering the number of imino protons in the i-motif moiety (6 $C{\cdot}C^+$ base pairs), the duplex (10 WC base pairs) and the G4 (3 tetrads x 4 iminos = 12), the G-rich strand is mainly folded (**Figure 4.19C**), since the ratio of signal intensities between Watson-Crick and G4 imino signals is very close to the one expected for scenario A (1:1.08 *versus* 1:1.2). The situation for the C-strand is less clear since the overall intensity of i-motif imino signals is lower than the expected ratio (0.6:1, corresponding to 6 $C{\cdot}C^+$ base pairs). These imino protons may be affected by exchange with water to a greater extent than WC and G4 imino protons. It can be observed from the 2D NMR of 35C2 that most cytosine H5 protons exhibit cross-peaks with amino protons in the 8.5 to 9.5 ppm region and with imino protons around 15.5 ppm, indicating that they are involved in $C{\cdot}C^+$ base pairs (**Figure 4.21A**). Additionally, the number and intensities of cytosine H5-H6 cross-peaks did not change upon complex formation; such a change would be expected if a significant population of unfolded C-rich strand was present (**Figure 4.21B**).

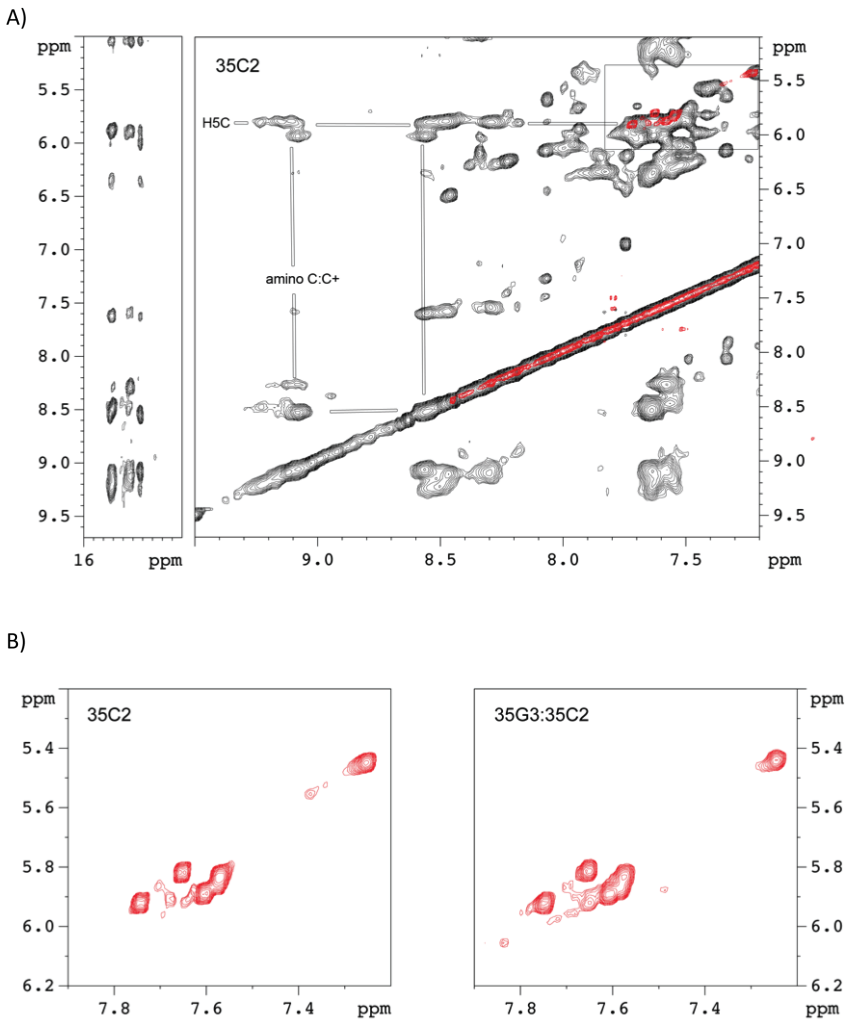


Figure 4.21: A) Two regions of the NOESY spectra of 35C2 (mixing time = 100 ms). Cross-peaks involving imino protons of C·C$^+$ base pairs are shown in the top left panel. Intraresidual cytosine amino/amino and amino/H5 cross-peaks are indicated in the top right panel. The overlaid TOCSY spectrum is shown in red. As expected for such a long and repetitive sequence, the poor signal dispersion complicates the sequential assignment. However, it can be observed that most cytosine H5 protons exhibit cross-peaks with amino protons in the 8.5 to 9.5 ppm region and with imino protons around 15.5 ppm, indicating that they are involved in C·C$^+$ base pairs. B) Enlarged H5-H6 region of the TOCSY spectra of 35C2 (left) and the 1:1 complex 35G3:35C2 (right). In the 1:1 complex, 35C2 and 35G3 were separately annealed and then mixed for two days before running the NMR. Buffer conditions: 20 mM KP$_i$ and 70 mM KCl at T = 5 °C, pH 7.0.

Although studies on promoter regions suggest that G4s and i-motifs are mutually exclusive when the G4 and i-motif are on opposite strands at the same location of a duplex,[40-43] the scenario here mirrors the telomeric region where the two structures are located at the duplex termini and where steric constraints are more relaxed.[62,63] Our studies now provide an opportunity to test whether G4s and i-motifs can co-exist next to each other in structures where

the G4/i-motif are flanked by two duplex overhangs, such as in DNA promoter regions. Given the thermodynamic stabilization and slower unfolding kinetics provided by the 2'F-ANA modification, it will be possible to study these putative intermediates without resorting to, for example, chemical footprinting and laser-tweezer-based single-molecule approaches. Most importantly, we can perform these experiments in a physiologically relevant buffer without the need to lower the pH. For instance, some reports investigating the mutually exclusive phenomenon stabilize i-motif and G4 structures at pH 5.5, 100 mM K^+; G4 separately at pH 7.4, 100 mM K^+; and i-motif alone at pH 5.5, 100 mM Li^+.

4.5 Conclusions

In conclusion, 2'F-ANA modifications slow down G4 and i-motif unfolding, thus providing an excellent tool to study G4/i-motif-duplex interconversion processes at neutral pH. When 2'F-ANA modified i-motif and G4 structures are pre-folded, their unfolding presents a significant barrier to duplex formation. This effect is more pronounced in the presence of G4/i-motif-duplex junctions, leading to very long-lasting intermediates in which i-motifs, G4s, and duplex structures may co-exist. While the presence of i-motif and G4 folds may be mutually exclusive under certain conditions, our results suggest that they may co-exist transiently as intermediates in telomeric sequences. Hybrid structures such as the ones studied here represent attractive tools for the discovery of selective ligands that can specifically stabilize i-motif structures, or target the G4/i-motif-duplex interface, which can in turn interfere with telomerase activity.[62,63] In this context, 2'F-ANA modification may become a useful method for building stable supramolecular DNA nanostructures based on the coexistence of i-motif and G4 complexes,[64] and for modulating the response of i-motif based nanodevices.[65,66] Interestingly, 2'F-araG parallel intramolecular G4 structures were found to be substrates of telomerase; however, the G-quadruplex structure in pre-folded G4 + i-motif (22G3 + 22C2) was not a substrate of telomerase. Therefore, the presence of the C-rich strand with the tendency to form a stable i-motif structure led to inhibition of telomerase. The mechanism of inhibition is currently under investigation. We believe these findings are significant as they may provide a step towards determining the role of the C-rich strand in telomere biology.

4.6 Experimental Data

4.6.1 Oligonucleotides Synthesis and Purification

Oligonucleotide synthesis was performed on an ABI 3400 DNA synthesizer (Applied Biosystems) at 1 μmol scale on Unylinker (Chemgenes) CPG solid support. Thymidine (dT), deoxycytidine (*N*-acetyl) (dC), and deoxyadenosine (*N*-Bz) (dA) phosphoramidites were used at 0.1 M concentration in acetonitrile and coupled for 110 s and deoxyguanosine (*N*-ibu) (dG) was coupled for 300 s. 0.15 M concentration was used for 2'F-araC which was coupled for 600 s. 2'F-araG and 2'F-riboG were used at 0.15 M concentration and coupled for 900 s. After completion of synthesis, the CPG was transferred to a 1.5 mL screw-cap eppendorf. Cleavage of the oligonucleotide from the CPG and removal of the nucleobase protecting groups were achieved by adding 1000 μL aqueous ammonium hydroxide for 16 h at 55 °C. The deprotection solution was centrifuged and decanted from the CPG. Samples were vented for 30 min, chilled in dry ice and evaporated to dryness. Crude oligonucleotides were purified by anion exchange HPLC on an Agilent 1200 Series instrument using a Protein-Pak DEAE 5PW column (7.5 x 75 mm) at a flow rate of 1 mL/min using a gradient of 0-24 % lithium perchlorate buffer (1.0 M) over 30 minutes at 60 °C. Under these conditions, the desired peaks eluted between 22 and 24 min. Samples were desalted on NAP-25 desalting columns according to manufacturer protocol. The extinction coefficients of 2'F-araC, 2'F-araG and 2'F-riboG were assumed to match those of the unmodified deoxynucleosides. Masses were verified by high resolution LC-MS.

Table 4.7: MS characterization of the studied sequences.

Code	Sequence (5'-3')	*Calculated mass (g/mole)*	*Mass found (g/mole) (-ve mode)*
22G0	A GGG TTA GGG TTA GGG TTA GGG	6963.1806	6963.2500
22G1	A (fG)GG TTA (fG)GG TTA (fG)(fG)G TTA (fG)GG	7053.1336	7053.1563
22G2	A (fG)GG TTA (fG)(fG)G TTA (fG)GG TTA (fG)GG	7053.1336	7052.9688
22G3	A (fG)GG TTA $(fG)_2$G TTA $(fG)_2$G TTA (fG)GG	7071.1242	7071.1641
22G4	A GGG TTA **G**GG TTA GGG TTA **G**GG	6999.1626	6998.5938
22G6	A GGG TTA **G**GG TTA GG**G** TTA **G**GG	7017.1536	7016.9688
22G7	A GGG TTA **GG**G TTA GG**G** TTA **G**GG	7035.1446	7035.0078
22C0	CCC TAA CCC TAA CCC TAA CCC T	6501.1306	6501.1563
22C2	$(fC)_3$ TAA $(fC)_3$ TAA $(fC)_3$ TAA $(fC)_3$ T	6717.0166	6717.0625
35G0	C ACA GAT GCG TTT A GGG TTA GGG TTA GGG TTA GGG	10978.0467	10978.0137
35G3	C ACA GAT GCG TTT A (fG)GG TTA (fG)(fG)G TTA (fG)(fG)G TTA (fG)GG	11085.9897	11085.0513
35C0	CCC TAA CCC TAA CCC TAA CCC T TTT CGC ATC TGT G	10497.7575	10496.9282
35C2	$(fC)_3$ TAA $(fC)_3$ TAA $(fC)_3$ TAA$(fC)_3$ T TTT CGC ATC TGT G	10713.6435	10712.8464

4.6.2 Thermal Melting Experiments

UV thermal denaturation data were obtained on a Varian CARY 100 UV-visible spectrophotometer equipped with a Peltier temperature controller. The concentration of oligonucleotides used was 10 µM for the G-rich single strands, 4 µM for the C-rich sequences. The total concentration for the duplexes was 2 µM. Samples were dissolved in appropriate buffer as indicated in the text. Concentrations were determined after quantitating the samples by UV absorbance at λ = 260 nm. Samples were heated to 90 °C for 10 minutes, then cooled slowly to room temperature, and stored at 5 °C for at least 18 h before the measurements were performed. Denaturation curves were acquired at 295 nm and 0.2 °C/min for G4s, 265 nm for i-motifs, and at 260 nm for the duplexes at a rate of 0.5 °C/min. In order to avoid water condensation on the cuvettes at low temperatures, samples were kept under a stream of dry nitrogen at temperatures below 12 °C. The dissociation temperatures were calculated as the midpoint of the transition (T_m) values using the first derivatives of the experimental data.

4.6.3 Circular Dichroism

Circular Dichroism (CD) studies were performed on a JASCO J-810 spectropolarimeter using a 1 mm path length cuvette. Temperature was maintained using the Peltier unit within the instrument. Spectra were recorded from 350-230 nm at a scan rate of 100 nm min^{-1} and a response time of 2.0 s with three acquisitions recorded for each spectrum. The spectra were normalized by subtraction of the background scan with buffer. Data were smoothed using the means-movement function within the JASCO graphing software. Oligonucleotide solutions for CD measurements were prepared with 20 mM KP_i + 70 mM KCl buffer (pH 5.8, 6.2, 6.6, 7.0 and 7.4) in a similar manner to that used for UV melting. The concentration of oligonucleotides used was 10 μM for the G-rich sequences at 25 °C, 20 μM for the C-rich sequences at 5 °C, and 15 μM for the duplexes at 5 °C. CD melting temperature experiments were performed on an Applied Photophysics CD spectrophotometer using a 1 mm path length cuvette. Spectra were recorded from 350-230 nm at a heating rate of 0.5 °C/min. Oligonucleotide solutions for CD melting experiments were either 40 μM or 100 μM in 20 mM KP_i and 70 mM KCl buffer pH 7.0.

4.6.4 NMR Kinetic Experiments

Samples for NMR experiments were dissolved in 9:1 H_2O/D_2O in 20 mM KP_i and 70 mM KCl buffer, pH 7.0. The sample concentration for 1D NMR kinetic experiments was 0.1 mM. The samples were annealed before running the NMR experiments as indicated in the main text. Spectra were acquired on Bruker Avance spectrometers operating at 600 MHz equipped with a cryoprobe. Two-dimensional spectra were recorded at 0.5 mM strand concentration in the same buffer. NOESY spectra were acquired with mixing times of 50 and 100 ms. TOCSY spectra were recorded with the standard MLEV-17 spin-lock sequence and a mixing time of 80 ms. Water suppression was achieved by including a WATERGATE module in the pulse sequence prior to acquisition.

4.6.5 Gel Electrophoresis Experiments

Native gel electrophoresis was performed utilizing 24 % polyacrylamide and TAE (Tris, Acetic acid, EDTA) buffer, pH 7.0. Running buffer was 1 x TAE. The gels were run at 280 V and 10 °C for 4 hours and 30 minutes. Annealed samples were heated to 90 °C for 10 minutes and then slowly cooled down to room temperature. For the pre-folded-mixed samples, 22G- and

22C-rich single strands were annealed separately and then mixed in equimolar amounts without further annealing. Gels were stained using SYBR Gold and visualized using Bio-Rad Gel Doc XR and controlled with the Image Lab software package.

4.6.6 Telomerase Activity Assays

Step 1: Sample preparation: 300 μL of 10 μM DNA solutions (22G0, 22C0, 22G3, and 22C2) were prepared in 20 mM KP_i and 70 mM KCl, pH 7.0. 22G-rich and 22C-rich samples were heated separately for 10 min at 90 °C. The metal block was then removed from the heater and the samples left to equilibrate to room temperature. Then samples were stored at 4°C overnight. The pre-folded 22G- and 22C-rich samples were mixed immediately before running the telomerase activity assays.

Step 2: Telomerase direct activity assay

The following reaction mixture was prepared to obtain 20 μL per sample: 2 μM of specified oligonucleotide sequences, 50 mM Tris-HCl pH 7.5, 1 mM $MgCl_2$, 5 mM DTT, 1 mM spermidine-HCl, 0.5 mM dATP, 0.5 mM dTTP, 5 μM dGTP, 0.33 μM [α-^{32}P] dGTP, 150 mM KCl, 2 fmol immunopurified telomerase and milliQ water. DNA was added into separate Protein LoBind Eppendorf tubes = 2 μl or 4 μl for extension with G and C rich DNA. Telomerase was pre-bound to DNA by incubating at 37 or 25 °C for 30 min. The reaction was initiated by adding 5 μL dNTP to the mix and then incubated at 37 °C for 1 h. The reaction was quenched with 80 μl 0.1 M EDTA with 2000 (cpm) 12-nt loading control. Samples were placed on ice for 5 mins then 33 μL 10 M ammonium acetate, 5 μL glycoblue (Ambion), and 330 μL ice-cold 100% ethanol were added to each tube. The samples were vortexed and incubated at -20 °C overnight. Samples were then centrifuged 13,000 g, 0-4 °C for at least 20 min. All ethanol was removed, and the pellet air-dried for 10 min. The pellet was resuspended in 10 μl formamide loading dye (90 % deionized formamide, 0.1 % Bromophenol Blue, 0.1 % Xylene Cyanol, and 10% 10x TBE).

Samples were heated for 10 min at 90 °C, placed on ice for at least 5 min, centrifuged at 13,000 g for 2 min and 3-5 μl of each sample loaded on a 12 % polyacrylamide sequencing gel. The gel was transferred onto a piece of Whatmann filter paper, covered with cling wrap and

dried for at least 20 min using a gel dryer with vacuum suction. The gel was exposed to a PhosphorImager screen and visualized the next day using a Typhoon scanner.

4.7 References

1. Blackburn, E. H.; Epel, E. S.; Lin, J., Human telomere biology: A contributory and interactive factor in aging, disease risks, and protection. *Science* **2015,** *350* (6265), 1193-8.
2. Sen, D.; Gilbert, W., Formation of Parallel 4-Stranded Complexes by Guanine-Rich Motifs in DNA and Its Implications for Meiosis. *Nature* **1988,** *334* (6180), 364-366.
3. Williamson, J. R.; Raghuraman, M. K.; Cech, T. R., Mono-Valent Cation Induced Structure of Telomeric DNA - the G-Quartet Model. *Cell* **1989,** *59* (5), 871-880.
4. Mergny, J. L.; De Cian, A.; Ghelab, A.; Sacca, B.; Lacroix, L., Kinetics of tetramolecular quadruplexes. *Nucleic Acids Res.* **2005,** *33* (1), 81-94.
5. Lipps, H. J.; Rhodes, D., G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol.* **2009,** *19* (8), 414-422.
6. Hansel-Hertsch, R.; Di Antonio, M.; Balasubramanian, S., DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nat. Rev. Mol. Cell Biol.* **2017,** *18* (5), 279-284.
7. Mergny, J. L.; Helene, C., G-quadruplex DNA: A target for drug design. *Nat. Med.* **1998,** *4* (12), 1366-1367.
8. Neidle, S.; Read, M. A., G-quadruplexes as therapeutic targets. *Biopolymers* **2001,** *56* (3), 195-208.
9. De Cian, A.; Lacroix, L.; Douarre, C.; Temime-Smaali, N.; Trentesaux, C.; Riou, J. F.; Mergny, J. L., Targeting telomeres and telomerase. *Biochimie* **2008,** *90* (1), 131-155.
10. 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.
11. De Cian, A.; Cristofari, G.; Reichenbach, P.; De Lemos, E.; Monchaud, D.; Teulade-Fichou, M. P.; Shin-Ya, K.; Lacroix, L.; Lingner, J.; Mergny, J. L., Reevaluation of telomerase inhibition by quadruplex ligands and their mechanisms of action. *Proc. Natl. Acad. Sci. USA* **2007,** *104* (44), 17347-52.
12. Zhao, Y.; Sfeir, A. J.; Zou, Y.; Buseman, C. M.; Chow, T. T.; Shay, J. W.; Wright, W. E., Telomere extension occurs at most chromosome ends and is uncoupled from fill-in in human cancer cells. *Cell* **2009,** *138* (3), 463-75.
13. Moye, A. L.; Porter, K. C.; Cohen, S. B.; Phan, T.; Zyner, K. G.; Sasaki, N.; Lovrecz, G. O.; Beck, J. L.; Bryan, T. M., Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nat. Commun.* **2015,** *6*, 1-12.
14. Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S., Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* **2013,** *5* (3), 182-6.

15. 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.
16. Laguerre, A.; Wong, J. M. Y.; Monchaud, D., Direct visualization of both DNA and RNA quadruplexes in human cells via an uncommon spectroscopic method. *Sci. Rep.* **2016,** *6*, 32141.
17. Fleming, A. M.; Ding, Y.; Alenko, A.; Burrows, C. J., Zika Virus Genomic RNA Possesses Conserved G-Quadruplexes Characteristic of the Flaviviridae Family. *ACS Infect. Dis.* **2016,** *2* (10), 674-681.
18. Gehring, K.; Leroy, J. L.; Gueron, M., A Tetrameric DNA-Structure with Protonated Cytosine.Cytosine Base-Pairs. *Nature* **1993,** *363* (6429), 561-565.
19. Berger, I.; Egli, M.; Rich, A., Inter-strand C-H···O hydrogen bonds stabilizing four-stranded intercalated molecules: Stereoelectronic effects of 04' in cytosine-rich DNA. *Proc. Natl. Acad. Sci. USA* **1996,** *93* (22), 12116-12121.
20. 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), 10739-41.
21. 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. USA* **2006,** *103* (52), 19658-63.
22. Day, H. A.; Huguin, C.; Waller, Z. A. E., Silver cations fold i-motif at neutral pH. *Chem. Commun.* **2013,** *49* (70), 7696-7698.
23. 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.
24. Abou Assi, H.; Harkness, R. W. V; Martin-Pintado, N.; Wilds, C. J.; Campos-Olivas, R.; Mittermaier, A. K.; Gonzalez, C.; Damha, M. J., Stabilization of i-motif structures by 2'-beta-fluorination of DNA. *Nucleic Acids Res.* **2016,** *44* (11), 4998-5009.
25. Simonsson, T.; Pribylova, M.; Vorlickova, M., A nuclease hypersensitive element in the human c-myc promoter adopts several distinct i-tetraplex structures. *Biochem. Biophys. Res. Commun.* **2000,** *278* (1), 158-66.
26. 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.
27. 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. *J. Am. Chem. Soc.* **2014,** *136* (11), 4172-4185.

28. Guo, K.; Gokhale, V.; Hurley, L. H.; Sun, D., Intramolecularly folded G-quadruplex and i-motif structures in the proximal promoter of the vascular endothelial growth factor gene. *Nucleic Acids Res.* **2008,** *36* (14), 4598-608.
29. Guo, K.; Pourpak, A.; Beetz-Rogers, K.; Gokhale, V.; Sun, D.; Hurley, L. H., Formation of pseudosymmetrical G-quadruplex and i-motif structures in the proximal promoter region of the RET oncogene. *J. Am. Chem. Soc.* **2007,** *129* (33), 10220-8.
30. Phan, A. T.; Gueron, 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.
31. Garavis, M.; Escaja, N.; Gabelica, V.; Villasante, A.; Gonzalez, C., Centromeric Alpha-Satellite DNA Adopts Dimeric i-Motif Structures Capped by AT Hoogsteen Base Pairs. *Chem. Eur. J.* **2015,** *21* (27), 9816-24.
32. Garavis, M.; Mendez-Lago, M.; Gabelica, V.; Whitehead, S. L.; Gonzalez, 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*, 13307.
33. 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.
34. 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.
35. Takahashi, S.; Brazier, J. A.; Sugimoto, N., Topological impact of noncanonical DNA structures on Klenow fragment of DNA polymerase. *Proc. Natl. Acad. Sci. USA* **2017,** *114* (36), 9605-9610.
36. Morin, G. B., The human telomere terminal transferase enzyme is a ribonucleoprotein that synthesizes TTAGGG repeats. *Cell* **1989,** *59* (3), 521-9.
37. Lingner, J.; Cech, T. R., Purification of telomerase from Euplotes aediculatus: Requirement of a primer 3' overhang. *Proc. Natl. Acad. Sci. USA* **1996,** *93* (20), 10712-10717.
38. Wang, H.; Blackburn, E. H., De novo telomere addition by Tetrahymena telomerase in vitro. *EMBO J.* **1997,** *16* (4), 866-79.
39. Cui, Y. X.; Kong, D. M.; Ghimire, C.; Xu, C. X.; Mao, H. B., Mutually Exclusive Formation of G-Quadruplex and i-Motif Is a General Phenomenon Governed by Steric Hindrance in Duplex DNA. *Biochemistry* **2016,** *55* (15), 2291-2299.
40. Dhakal, S.; Yu, Z. B.; Konik, R.; Cui, Y. X.; Koirala, D.; Mao, H. B., G-Quadruplex and i-Motif Are Mutually Exclusive in ILPR Double-Stranded DNA. *Biophys. J.* **2012,** *102* (11), 2575-2584.
41. Sutherland, C.; Cui, Y. X.; Mao, H. B.; 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.

42. 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.
43. Brown, R. V.; Wang, T.; Chappeta, V. R.; Wu, G.; Onel, B.; Chawla, R.; Quijada, H.; Camp, S. M.; Chiang, E. T.; Lassiter, Q. R.; Lee, C.; Phanse, S.; Turnidge, M. A.; Zhao, P.; Garcia, J. G. N.; Gokhale, V.; Yang, D.; Hurley, L. H., The Consequences of Overlapping G-Quadruplexes and i-Motifs in the Platelet-Derived Growth Factor Receptor beta Core Promoter Nuclease Hypersensitive Element Can Explain the Unexpected Effects of Mutations and Provide Opportunities for Selective Targeting of Both Structures by Small Molecules To Downregulate Gene Expression. *J. Am. Chem. Soc.* **2017,** *139* (22), 7456-7475.
44. Phan, A. T.; Mergny, J. L., Human telomeric DNA: G-quadruplex, i-motif and watson-crick double helix. *Nucleic Acids Res.* **2002,** *30* (21), 4618-4625.
45. Li, W.; Wu, P.; Ohmichi, T.; Sugimoto, N., Characterization and thermodynamic properties of quadruplex/duplex competition. *Febs Lett.* **2002,** *526* (1-3), 77-81.
46. Bucek, P.; Jaumot, J.; Avino, A.; Eritja, R.; Gargallo, R., pH-Modulated Watson-Crick Duplex-Quadruplex Equilibria of Guanine-Rich and Cytosine-Rich DNA Sequences 140 Base Pairs Upstream of the c-kit Transcription Initiation Site. *Chem. Eur. J.* **2009,** *15* (46), 12663-12671.
47. Li, W.; Miyoshi, D.; Nakano, S.; Sugimoto, N., Structural competition involving G-quadruplex DNA and its complement. *Biochemistry* **2003,** *42* (40), 11736-44.
48. Kumar, N.; Sahoo, B.; Varun, K. A.; Maiti, S.; Maiti, S., Effect of loop length variation on quadruplex-Watson Crick duplex competition. *Nucleic Acids Res.* **2008,** *36* (13), 4433-42.
49. Peng, C. G.; Damha, M. J., G-quadruplex induced stabilization by 2'-deoxy-2'-fluoro-D-arabinonucleic acids (2'F-ANA). *Nucleic Acids Res.* **2007,** *35* (15), 4977-88.
50. Lech, C. J.; Li, Z.; Heddi, B.; Phan, A. T., 2'-F-ANA-guanosine and 2'-F-guanosine as powerful tools for structural manipulation of G-quadruplexes. *Chem. Commun.* **2012,** *48* (93), 11425-7.
51. Martin-Pintado, N.; Yahyaee-Anzahaee, M.; Deleavey, G. F.; Portella, G.; Orozco, M.; Damha, M. J.; Gonzalez, C., Dramatic effect of furanose C2' substitution on structure and stability: directing the folding of the human telomeric quadruplex with a single fluorine atom. *J. Am. Chem. Soc.* **2013,** *135* (14), 5344-7.
52. Phan, A. T.; Gueron, M.; Leroy, J. L., Investigation of unusual DNA motifs. *Method Enzymol.* **2001,** *338*, 341-371.
53. Mergny, J. L.; Phan, A. T.; Lacroix, L., Following G-quartet formation by UV-spectroscopy. *Febs Lett.* **1998,** *435* (1), 74-78.

54. Leroy, J. L.; Gueron, M.; Mergny, J. L.; Helene, C., Intramolecular folding of a fragment of the cytosine-rich strand of telomeric DNA into an i-motif. *Nucleic Acids Res.* **1994,** *22* (9), 1600-6.
55. Kanehara, H.; Mizuguchi, M.; Tajima, K.; Kanaori, K.; Makino, K., Spectroscopic evidence for the formation of four-stranded solution structure of oligodeoxycytidine phosphorothioate. *Biochemistry* **1997,** *36* (7), 1790-1797.
56. Kypr, J.; Kejnovska, I.; Renciuk, D.; Vorlickova, M., Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.* **2009,** *37* (6), 1713-1725.
57. Martin-Pintado, N.; Yahyaee-Anzahaee, M.; Campos-Olivas, R.; Noronha, A. M.; Wilds, C. J.; Damha, M. J.; Gonzalez, C., The solution structure of double helical arabino nucleic acids (ANA and 2 ' F-ANA): effect of arabinoses in duplex-hairpin interconversion. *Nucleic Acids Res.* **2012,** *40* (18), 9329-9339.
58. 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.
59. Oganesian, L.; Moon, I. K.; Bryan, T. M.; Jarstfer, M. B., Extension of G-quadruplex DNA by ciliate telomerase. *EMBO J.* **2006,** *25* (5), 1148-59.
60. Zaug, A. J.; Podell, E. R.; Cech, T. R., Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proc. Natl. Acad. Sci. USA* **2005,** *102* (31), 10864-9.
61. Konig, S. L.; Huppert, J. L.; Sigel, R. K.; Evans, A. C., Distance-dependent duplex DNA destabilization proximal to G-quadruplex/i-motif sequences. *Nucleic Acids Res.* **2013,** *41* (15), 7453-61.
62. Ren, J.; Qu, X.; Trent, J. O.; Chaires, J. B., Tiny telomere DNA. *Nucleic Acids Res.* **2002,** *30* (11), 2307-15.
63. Krauss, I. R.; Ramaswamy, S.; Neidle, S.; Haider, S.; Parkinson, G. N., Structural Insights into the Quadruplex-Duplex 3 ' Interface Formed from a Telomeric Repeat: A Potential Molecular Target. *J. Am. Chem. Soc.* **2016,** *138* (4), 1226-1233.
64. Cao, Y.; Gao, S.; Yan, Y.; Bruist, M. F.; Wang, B.; Guo, X., Assembly of supramolecular DNA complexes containing both G-quadruplexes and i-motifs by enhancing the G-repeat-bearing capacity of i-motifs. *Nucleic Acids Res.* **2016**.
65. Dong, Y. C.; Yang, Z. Q.; Liu, D. S., DNA Nanotechnology Based on i-Motif Structures. *Acc. Chem. Res.* **2014,** *47* (6), 1853-1860.
66. Yatsunyk, L. A.; Mendoza, O.; Mergny, J. L., "Nano-oddities": Unusual Nucleic Acid Assemblies for DNA-Based Nanostructures and Nanodevices. *Acc. Chem. Res.* **2014,** *47* (6), 1836-1844.

Chapter 5

Contributions to Knowledge and Future Work

"You measure the size of the accomplishment by the obstacles you had to overcome to reach your goals".
–Booker T. Washington

5.1 Summary and Contributions to Knowledge

5.1.1 Stabilization of i-Motif Structures by 2'-β-Fluorination of DNA (Chapter 2)

Our work demonstrated that incorporating 2'F-araC modifications in i-motif structures leads to significant stabilization over a wide pH range, an exciting result given that very few chemical modifications are tolerated in the i-motif structure. Furthermore, the 2'F-araC modification stabilized intermolecular centromeric and intramolecular telomeric sequences in a position-independent manner. Despite the fact that 2'F-araC exhibits a lower pK_a compared to deoxycytidine (3.9 *versus* 4.4, respectively), the change in pH of mid transition ($pH_{1/2}$) in the modified i-motif structures was remarkable (+0.7 for centromeric sequences and +0.8 for telomeric sequences), allowing for the observation of these structures at neutral pH. NMR structural determination revealed that the 2'F-araC residues adopt a C2'-*endo* sugar pucker, instead of the C3'-*endo* conformation that is usually found in unmodified structures, with the fluorine atom oriented in the major groove. Therefore, 2'F-araC modifications do not perturb the hydrogen-bonding network that holds the structure together, but instead lead to additional sequential and inter-strand electrostatic interactions that are absent in the unmodified structure. The above results allow the utilization of i-motif structures in several applications, most importantly in biological assays that require physiological temperature and pH conditions.

5.1.2 Probing Synergistic Effects of DNA Methylation and 2'-β-Fluorination on i-Motif Stability (Chapter 3)

As a follow-up on the results reported in Chapter 2, our aim in Chapter 3 was to search for a modification or combination of modifications that would lead to even higher thermal stability at physiological pH along with higher $pH_{1/2}$ values compared to the 2'F-araC modification. For this purpose, we synthesized a novel cytidine analogue, 2'-deoxy-2'-fluoro-5-methyl-arabinocytidine (5-Me-2'F-araC), that combines the two most stabilizing i-motif modifications, 5-methylcytosine nucleobase and 2-fluoroarabinose sugar in the same nucleotide. This modification was found to exhibit a similar stabilization effect to the 2'F-araC modification. Next, we investigated a combination of different cytidine nucleotides and their impact on i-motif structures. The results of this work highlighted the significance of the base-pairing nature on i-motif stability. For instance, a structure stabilized by 2'F-araC/5-Me-dC base pairs (HJ-2) exhibits a $pH_{1/2}$ value of 7.17 compared to 6.53 for a structure containing 2'F-araC/2'F-araC and

5-Me-dC/5-Me-dC base pairs (HJ-1). This suggests that it is possible to tune the pH and thermal melting response of i-motif structures by carefully selecting the position and type of modifications.

5.1.3 2'-Fluoroarabinonucleic Acid Modification Traps G-Quadruplex and i-Motif Structures in Human Telomeric DNA (Chapter 4)

Taking advantage of the high stabilization observed with 2'F-araC modifications, the influence of 2'F-araG and 2'F-araC modifications on G4/i-motif-duplex interconversion kinetics was investigated at neutral conditions. For this purpose, we first stabilized telomeric G4 structures and determined that 2'F-araG modifications stabilize the desired parallel intramolecular G4 structure exclusively over competing conformers. These modified G4 structures were found to be substrates of telomerase and were elongated in the typical +6nt pattern. This result confirmed that the parallel topology is a requirement for recognition of G4 structures by telomerase.

Our ability to stabilize the telomeric G4 and i-motif structures at opposite strands of a duplex made it possible to study the kinetics of duplex formation. When 2'F-araC modified i-motif and 2'F-araG modified G4 structures were pre-folded, their unfolding poses a significant barrier to duplex formation. This leads to very long-lasting intermediates in which i-motifs, G4s, and duplex structures co-exist at neutral conditions. These results have important implications in telomere biology since “trapping” both G4 and i-motif structures make it possible to examine how i-motifs affect telomerase activity under neutral conditions. Indeed, preliminary telomerase activity assays showed that in the presence of an unfolded C-rich strand or an i-motif structure, telomerase does not efficiently elongate parallel G4 structures. The mechanisms of G4 elongation and the inhibition of telomerase by C-rich sequences are currently under investigation in our laboratory.

5.1.4 Synthesis, Structure, and Conformational Analysis of Nucleoside Analogues Comprising Six-Membered 1,3-Oxathiane Sugar Rings (Appendix I)

Given the scarce examples of six-membered ring nucleosides, we describe in Appendix I the synthesis of novel nucleoside analogues comprising 1,3-oxathiane sugar rings. The pyrimidine and purine analogues were synthesized from readily available starting materials *via* a

ring-closing pathway in good yields. From the X-ray structure obtained for the N^7-β-OMe-purine analogue, we were able to unambiguously determine that the six-membered sugar adopts a perfect chair conformation with the substituents placed in equatorial positions. Despite some structural similarity with their five-membered counterpart (3TC), no significant anti-HIV activity was detected for the 1,3-oxathiane nucleosides.

5.2 Suggestions for Future Work

The results reported in this thesis bring i-motif structures into focus in several areas of research. For that purpose, several follow-up experiments can be performed in efforts to clarify the structural and functional properties of the modified i-motifs. In Chapter 3, we determined that HJ-2 (C**C^{m}C^{m}** TAA C**fCfC** TAA C**fCfC** TAA C**C^{m}C^{m}**) exhibits the highest thermal stability at pH 7.4 (32.2 °C). This stability, based on NMR analysis, was hypothesized to be a result of 2'F-araC/5-Me-dC base-pairing leading to several favorable electrostatic interactions. The next steps would be to carry out 2D NMR analysis on HJ-3 (C**fC^{m}fC^{m}** TAA C**fCfC** TAA C**fCfC** TAA C**fC^{m}fC^{m}**, $T_{1/2}$ = 29.5 °C) since it exhibits a $pH_{1/2}$ of 7.34. Likewise, comparison of HJ-3 to HJ-1 (C**fCC^{m}** TAA C**C^{m}fC** TAA C**fCC^{m}** TAA C**C^{m}fC**, $pH_{1/2}$ = 6.53) would help us understand why these i-motifs exhibit very different $pH_{1/2}$ values. Comparison of HJ-1 ($pH_{1/2}$ = 6.53) with HJ-2 ($pH_{1/2}$ = 7.17) which have the same modifications would provide further information on how positional effects affect i-motif stability.

Studies from the Bryan group (Children's Medical Research Institute, Sydney) have demonstrated that telomerase is able to elongate both ciliate and human telomeric sequences when they are pre-folded into *intermolecular* parallel G-quadruplex structures, but the same is not true for the antiparallel G4s. In Chapter 4, we demonstrated the ability of telomerase to extend parallel *intramolecular* 2'F-araG modified G4 structures. From these results, we deduce that telomerase interacts differently with diverse G4 folding topologies, with the parallel conformation appearing to be a requirement for telomerase recognition and elongation. Since it has been suggested that, prior to elongation, telomerase partially unwinds the dimeric G4 structures, it is of utmost importance to determine the mechanism of intramolecular G4 elongation, especially given the slow unfolding kinetics and high thermal stability of our modified G4 structures (T_m = 80.2 °C). Preliminary experiments carried out in our laboratory in collaboration with the González group (CSIC, Madrid) have shown that the ^{19}F-NMR signals in

folded G4 structures are not visible, whereas they become visible upon melting. Hence, the unfolding of our pre-folded fluorinated G4s, upon addition of telomerase, could in principle be monitored *via* ^{19}F-NMR spectroscopy.

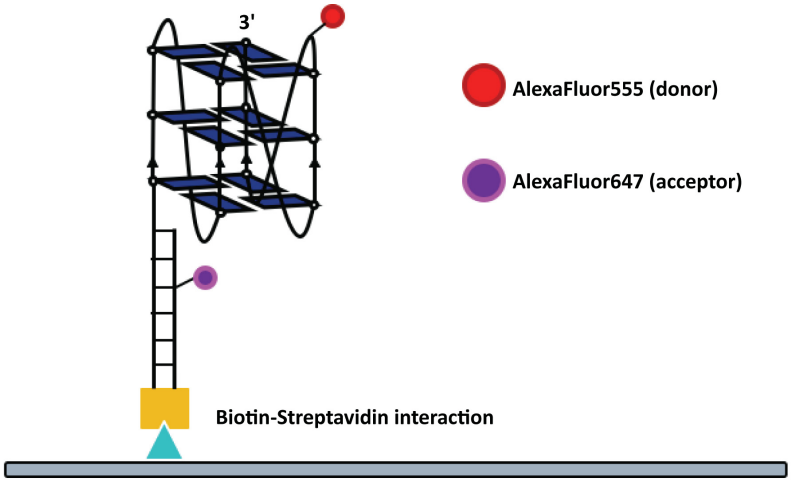


Figure 5.1: Proposed single molecule FRET experiments.

Another method to monitor telomerase activity is through single molecule FRET experiments. This work is in progress in our lab in collaboration with the Bryan group and the van Oijen group (University of Wollongong, Australia). As shown in **Figure 5.1**, a G4 structure with a duplex tail has been synthesized in order to position the dyes at a distance optimal for FRET analysis. AlexaFluor555 can be utilized as the donor and AlexaFluor647 as the acceptor. We anticipate that when telomerase is added to the system, partial reduction in FRET signal (due to partial unfolding) will be observed. When telomerase and nucleotides are added, we expect a rapid/complete unfolding of the G4 structures. These experiments will help determine the unfolding of the G4s on a single molecule basis. Together, the results of ^{19}F-NMR and single molecule FRET will provide significant knowledge on the mechanism of telomere elongation, which is of considerable interest given the role of telomerase activity in cellular aging and cancer.

Furthermore, several direct telomerase activity assays should be performed to determine whether telomerase is able to unwind telomeric G4 structures in the presence of an i-motif.

Table 5.1: Sequences to be utilized in telomerase activity assays.

Code	Sequence (5'-3')
22G0	A GGG TTA GGG TTA GGG TTA GGG
22G3	A (fG)GG TTA (fG)(fG)G TTA (fG)(fG)G TTA (fG)GG
22C0	CCC TAA CCC TAA CCC TAA CCC T
22C2	((fC)(fC)(fC) TAA)$_3$ (fC)(fC)(fC) T
35G0	C ACA GAT GCG TTT A GGG TTA GGG TTA GGG TTA GGG
35G3	C ACA GAT GCG TTT A (fG)GG TTA (fG)$_2$G TTA (fG)$_2$ G TTA (fG)GG
35C0	CCC TAA CCC TAA CCC TAA CCC T TTT CGC ATC TGT G
35C2	(fC)$_3$ TAA (fC)$_3$ TAA (fC)$_3$ TAA (fC)$_3$ T TTT CGC ATC TGT G

Telomerase activity assays can be performed for the following combinations of G- and C-rich strands both as annealed duplexes and as pre-folded mixtures:

- 22G3 + 22C2 (intramolecular parallel G4 + i-motif)
- 22G3 + 22C0 (intramolecular parallel G4 + single-stranded C-rich)
- 22G0 + 22C0 (G4 as a mixture of parallel and antiparallel + single-stranded C-rich)
- 22G0 + 22C2 (G4 as a mixture of parallel and antiparallel + i-motif)
- 35G3 + 35C2 (parallel G4 + i-motif)
- 35G3 + 35C0 (parallel G4 + single-stranded C-rich)
- 35G0 + 35C0 (G4 as a mixture of parallel and antiparallel + single-stranded C-rich)
- 35G0 + 35C2 (G4 as a mixture of parallel and antiparallel + i-motif)
- A DNA duplex with a G4 overhang, formed by mixing stoichiometric amounts of 35G3 and its partially complementary sequence 5'-AAA CGC ATC TGT G-3'.

Based on the results reported in Chapter 4, in which G4 and i-motif structures co-exist in the same telomeric complex, it will be interesting to incorporate our modifications in duplexes

mimicking gene promoter regions, where the G4 and i-motif structures are flanked by two DNA duplexes (**Figure 5.2A**). This will allow us to probe whether steric hindrance is the underlying cause of 'mutual exclusivity', as has been previously proposed. Importantly, we are now able to perform these experiments under physiological conditions. Moreover, since the telomeric G-rich 3'-overhangs contain multiple G-tracts, several consecutive G4 structures could form. Therefore, it will be interesting to perform telomerase activity assays on these more biologically-relevant scenarios in which multiple G4 structures on the G-rich strand exist in combination with an i-motif structure on the complementary strand (**Figure 5.2B** and **5.2C**).

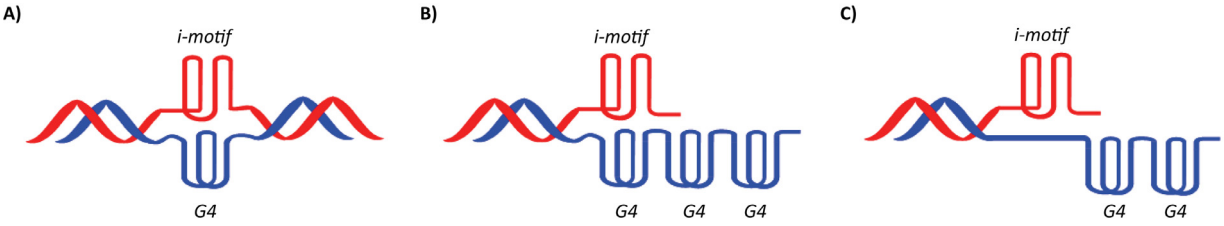


Figure 5.2: Proposed designs: A) G4 and i-motif structures flanked between two DNA duplexes in gene promoter regions. B and C) Telomeric G-rich strands possessing multiple G4 structures with an i-motif structure on the complementary strand.

5.3 List of Publications

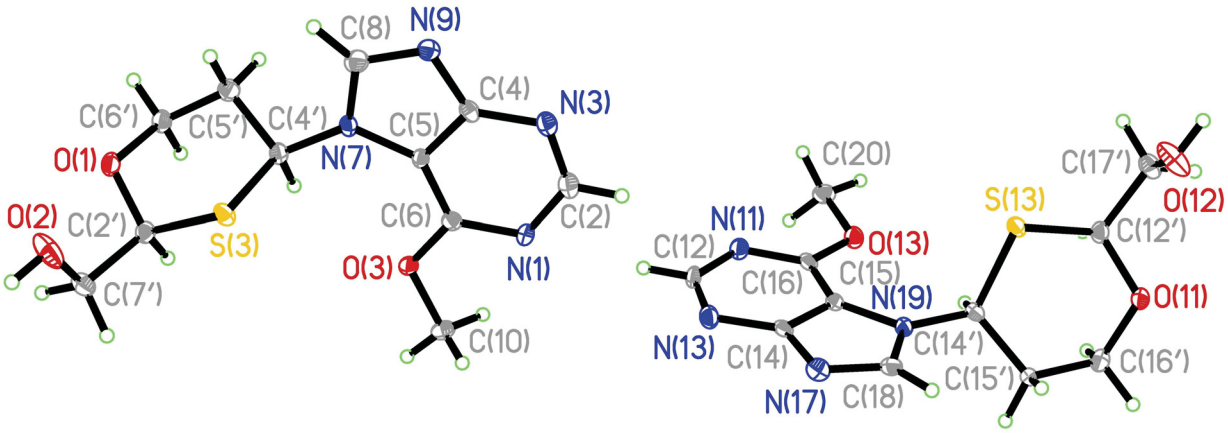
1. Abou Assi, H.; Lin, Y.C.; Serrano, I.; González, C.; Damha, M.J. Probing synergistic effects of DNA methylation and 2'-β-fluorination on i-motif stability, *Chemistry – A European Journal*, **2017**, doi: 10.1002/chem.201704591.
2. Abou Assi, H.; El-Khoury, R.; González, C.; Damha, M.J. 2'-Fluoroarabinonucleic acid modification traps G-quadruplex and i-motif structures in human telomeric DNA. *Nucleic Acids Research*, **2017**, *45*, 11535-11546.
3. Lietard, J.; Abou Assi, H.; Gomez-Pinto, I.; González, C.; Somoza. M.M.; Damha, M.J. Mapping the affinity landscape of thrombin-binding aptamers on 2'F-ANA/DNA chimeric G-quadruplex microarrays. *Nucleic Acids Research*, **2017**, *45,* 1619-1632.
4. Abou Assi, H.; 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 Research*, **2016**, *44*, 4998-5009.
5. Abou Assi, H.; Martinez-Montero, S.; Dixit, D.; Chua, Z.; Bohle, D.S.; Damha, M.J. Synthesis, Structure, and Conformational Analysis of Nucleoside Analogues Comprising Six-Membered 1,3-Oxathiane Sugar Rings. *European Journal of Organic Chem*istry, **2015**, 1945-1953.

5.4 List of Conference Presentations

1. Abou Assi, H.; González, C.; Damha, M.J. Trapping G-quadruplex/i-Motif Intermediates of Human Telomeric DNA, 6th International Meeting on Quadruplex Nucleic Acids, May 31st-June 3rd, 2017, Prague, Czech Republic (Poster presentation).
2. Abou Assi, H.; González, C.; Damha, M.J. Trapping G-quadruplex/i-Motif Intermediates of Human Telomeric DNA, 12th Annual Meeting of the Oligonucleotide Therapeutics Society, September 2016, Montreal, Qc, Canada (Poster presentation; received best poster presentation award).
3. Abou Assi, H.; Harkness, R.W.; Martın-Pintado, N.; Mittermaier, A.K.; González, C.; Damha, M.J. Stabilization of i-Motif Structures by 2'-β-Fluorination of DNA, XXII International Roundtable on Nucleosides, Nucleotides and Nucleic Acids (IRT), July 2016, Paris, France (Poster presentation; received IRT travel grant).
4. Abou Assi, H.; Harkness, R.W.; Martın-Pintado, N.; Campos-Olivas, R.; Mittermaier, A.K.; González, C.; Damha, M.J. Pronounced Effect of 2'F-arabinose (2'F-ANA) Substitution on the Conformation and Stability of I-motif Structures, 250th American Chemical Society National Meeting & Exhibition, August 16-20, 2015, Boston, MA, USA (Oral presentation).
5. Abou Assi, H.; Martín-Pintado, N.; González, C.; Damha, M.J. The Effect of 2'-Fluoroarabinose Modifications on the Formation and Stability of I-motifs (Cytosine Quadruplexes), 97th Canadian Chemistry Conference and Exhibition, June 2, 2014, Vancouver, Canada (Oral presentation; received best talk award).
6. Abou Assi, H.; Dixit, D.M.; Damha, M.J. Design, Synthesis and Biological Evaluation of Novel Heterocyclic Six-Membered Nucleoside Analogues, CIHR Drug Development and Training Program Annual Research Retreat, May 31, 2013, Montreal, Canada (Poster presentation; received best poster presentation award).
7. Abou Assi, H.; Dixit, D.M.; Damha, M.J. Total Synthesis and Biological Evaluation of Nucleoside Analogues Comprising Heterocyclic Six-Membered Ring Sugars, 96th Canadian Chemistry Conference and Exhibition, May 30, 2013, Quebec City, QC., Canada (Oral presentation).
8. Abou Assi, H.; Dixit, D.M.; Damha, M.J. Synthesis of Nucleoside Analogues Comprising Six-Membered Sugar Rings, DDTP Research Retreat, March 30, 2012, Montreal, QC., Canada (Oral and poster presentations).

Appendix I

Synthesis, Structure, and Conformational Analysis of Nucleoside Analogues Comprising Six-Membered 1,3-Oxathiane Sugar Rings



The majority of this chapter is reproduced from: "Synthesis, Structure, and Conformational Analysis of Nucleoside Analogues Comprising Six-Membered 1,3-Oxathiane Sugar Rings", **Hala Abou Assi**, Saul Martinez-Montero, Dilip Dixit, Zhijie Chua, D. Scott Bohle and Masad J. Damha, *European Journal of Organic Chemistry*, **2015**, 1945-1953.

"Develop success from failures. Discouragement and failure are two of the surest stepping stones to success."

–Dale Carnegie

AI.1 Introduction

The successful journey of nucleoside analogues as potent anticancer[1] and antiviral[2] therapeutics began several decades ago. The discovery of 3'-azido-2',3'-dideoxythymidine (AZT)[3] as a nucleoside reverse transcriptase inhibitor (NRTI) that prevents HIV replication initiated the search for novel 2',3'-dideoxynucleosides (ddNs) with enhanced activity and minimal toxicity such as Zalcitabine (ddC),[4] Didanosine (ddI),[4] Stavudine (d4T),[5] and the carbocyclic analogue Abacavir (ABC)[6] (**Figure AI.1**). These nucleoside derivatives are considered chain terminators since the absence of the 3'-OH prevents chain elongation and slows down viral replication once the corresponding triphosphate nucleosides are incorporated into the viral DNA by HIV reverse transcriptase.

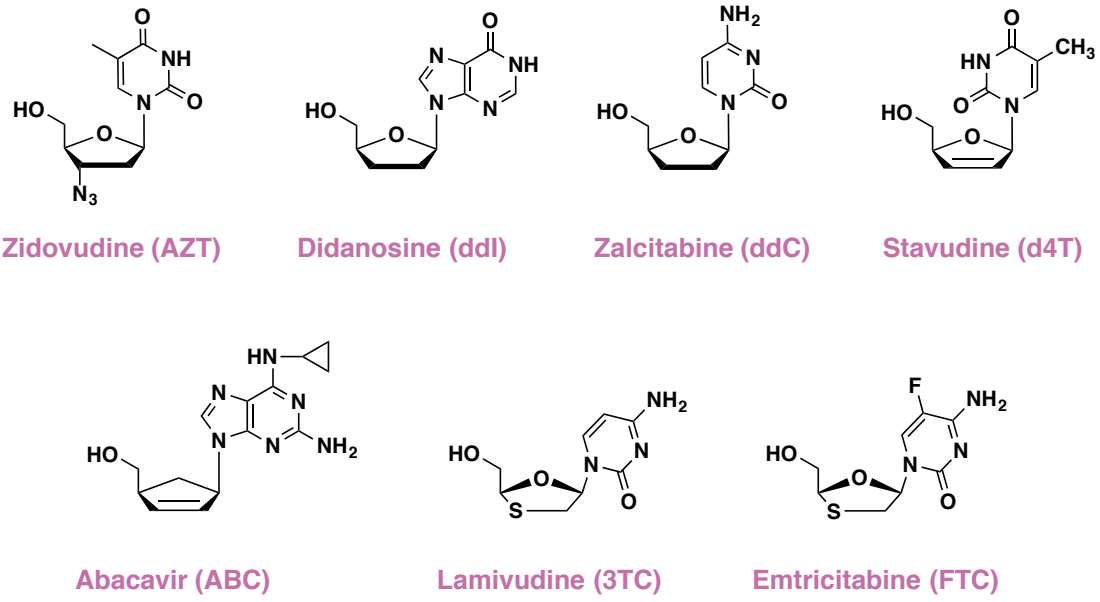


Figure AI.1: Clinically approved nucleoside reverse transcriptase inhibitors.

Five-membered 2',3'-dideoxy heteronucleosides wherein one or more heteroatom is introduced in the furanose ring have also received much attention due to their antiviral activity.[7] One of the early examples that shed light on the therapeutic potential of heterocyclic nucleosides was (-)-2',3'-dideoxy-3'-thiacytidine (3TC),[8] which was approved by the FDA in 1995 for combating HIV infection. Its 5-fluorocytosine derivative, Emtricitabine [(-)FTC], was also approved as an anti-HIV drug in 2003.[9] Despite the success associated with antiretroviral combination therapy,[2] the development of resistance resulting from the extensive use of existing

clinically approved compounds remains a major problem.[10,11] Therefore, the search for new potent anti-HIV agents is a very active area of research.

Few examples of ring-expanded nucleoside/nucleotide analogues have been reported in the literature.[12,13] *Cis*-substituted cyclohexenyl and cyclohexanyl nucleoside analogues have been synthesized and shown to have moderate activity against coxsackie and HSV-1 viruses (**Figure AI.2A**).[14] Nucleoside analogues with 1,4-heteroatom six-membered ring systems (*e.g.*, 1,4-dioxane, 1,4-oxathiolane and 1,4-oxazine) replacing the furanose moiety have also been reported (**Figure AI.2B**).[15-19] Six-membered 1,3-dioxane nucleoside derivatives having the nucleobase at position 5 (**Figure AI.2C**) have been reported to exhibit moderate anti-HIV activity.[20]

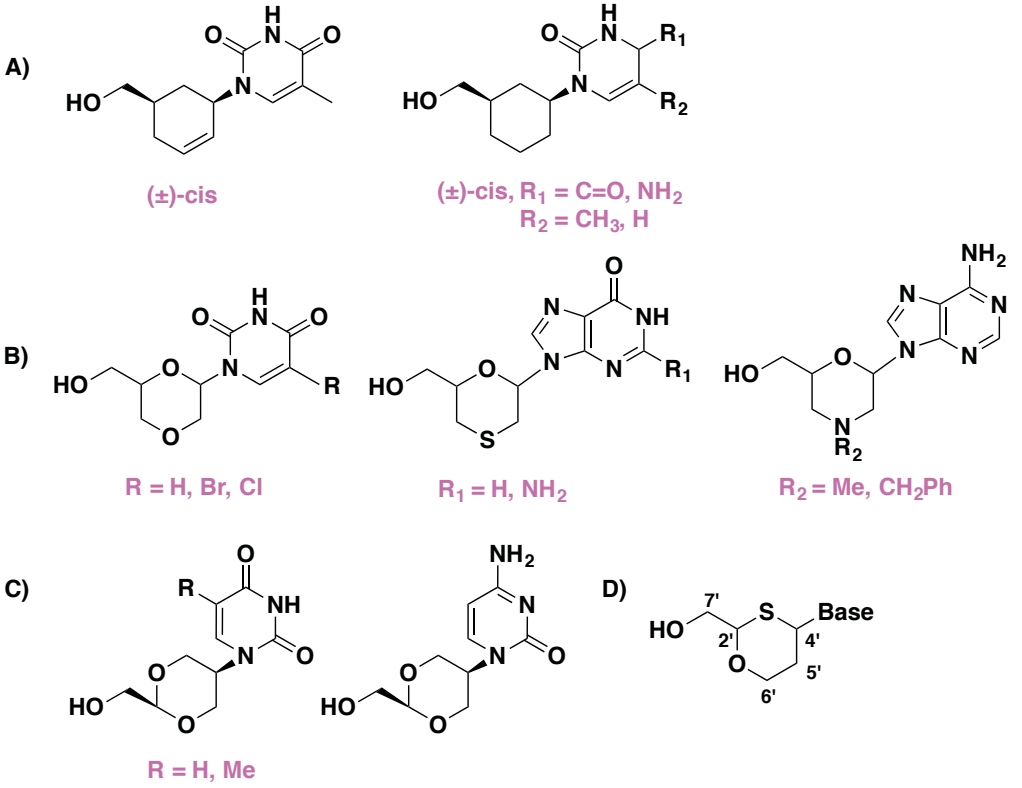


Figure AI.2: A, B and C show previously reported six-membered nucleoside analogues. D shows the general structure of the six-membered 1,3-oxathiane nucleoside analogues synthesized in this work.

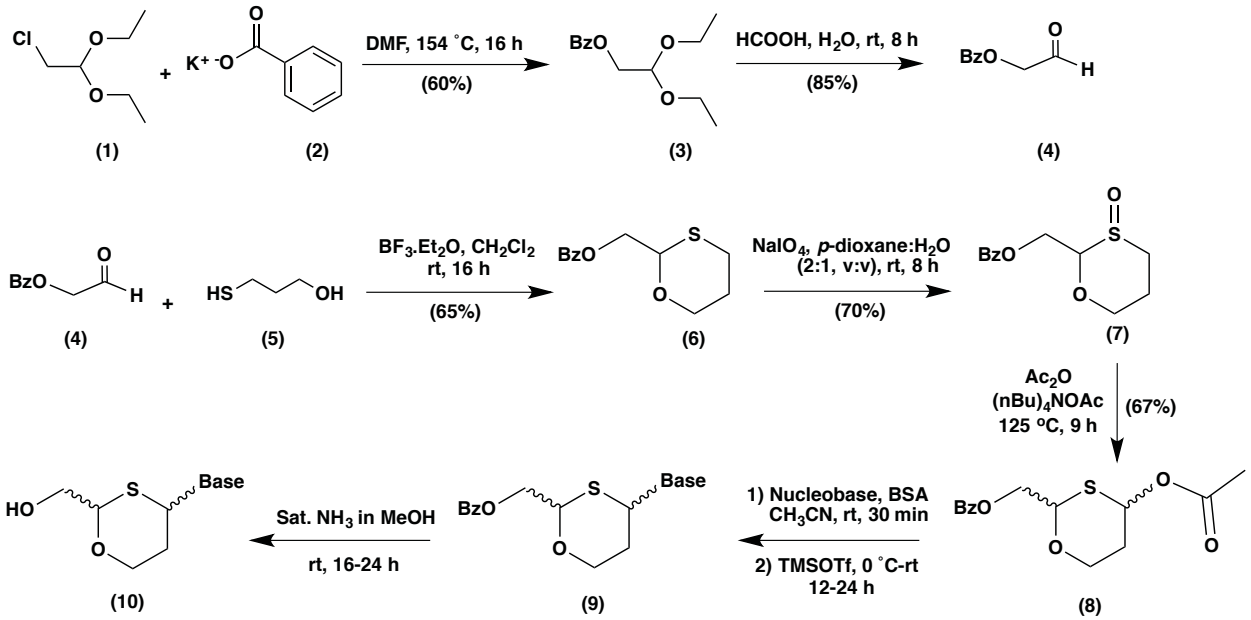
Given the scarce examples of ring-expanded nucleosides, and as part of our program on the synthesis and conformational analysis of heteronucleosides and nucleoside analogues with expanded ring sugars,[21-24] this appendix describes the synthesis of nucleoside analogues comprising six-membered 1,3-oxathiane sugar rings. The hypothesis we aimed to test was

whether the 1,3-configuration of the heteroatoms, with the base at the 4'-position, although more challenging to access than the previously reported 1,4-derivatives, may share some of the properties of the 1,3-furanoside nucleosides (*e.g.*, 3TC).

AI.2 Synthesis of 1,3-Oxathiane Nucleoside Analogues

The synthesis of 1,3-oxathiane nucleosides with various pyrimidine heterocyclic bases is outlined in **Scheme AI.1**. Benzoyloxyacetaldehyde (**4**) was prepared according to the procedure reported by Du and Watanabe[25] *via* treatment of commercially available 2-chloro-1,1-diethoxyethane (**1**) with potassium benzoate (**2**) in DMF in the presence of catalytic potassium iodide, followed by hydrolysis of the resulting acetal with aqueous formic acid. The key step of the synthesis was the condensation of benzoyloxyacetaldehyde with 3-mercapto-1-propanol in order to obtain the 1,3-oxathiane six-membered sugar moiety. We investigated several Lewis acid catalysts including $TiCl_4$,[26] $ZrCl_4$,[27] and $LiBF_4$;[28] however, the best result was obtained with $BF_3.Et_2O$ in dichloromethane under anhydrous conditions leading to a moderate yield of compound **6**. Oxidation of the ring sulfur to the corresponding sulfoxide was performed using two equivalents of sodium periodate. In order to transform the sulfoxide into a glycosylation precursor, a 4-acetoxy leaving group was introduced in the 1,3-oxathiane six-membered ring *via* Pummerer rearrangement of the sulfoxide **7** in refluxing acetic anhydride/tetrabutylammonium acetate[29] to afford **8** in good yield (69%).

Glycosylation of **8** under Vorbrüggen's conditions[30] with various pyrimidine heterocyclic bases produced α/β anomeric mixtures **9a-c** that were not separable by column chromatography. We next focused on finding procedures to resolve these diastereomeric mixtures. Towards this goal we found analytical reverse-phase HPLC (RP-HPLC) conditions to separate the anomers of the cytosine analogue **10a** (**Figure AI.3**). However, the resolution was lost when higher amounts were injected in order to collect enough material for biological testing and characterization (**Figure AI.3**).



	Glycosylation yield (%)	Deprotection yield (%)	α:β ratio (NMR)
a. B=cytosine	54	86	1.7:1
b. B=thymine	68	89	1.9:1
c. B=5F-cytosine	72	89	1.8:1

Scheme AI.1: Synthesis of pyrimidine 1,3-oxathiane nucleosides.

Glycosylation of **8** with silylated 6-chloropurine under Vorbrüggen's conditions resulted in N^9/N^7 isomeric mixtures. Four compounds were present in the reaction mixture since α and β anomers result from each glycosylated regioisomer. N^9 and N^7 isomers were easily separable by silica gel column chromatography, with the kinetic N^7 regioisomer obtained as the major product (**Scheme AI.2**).

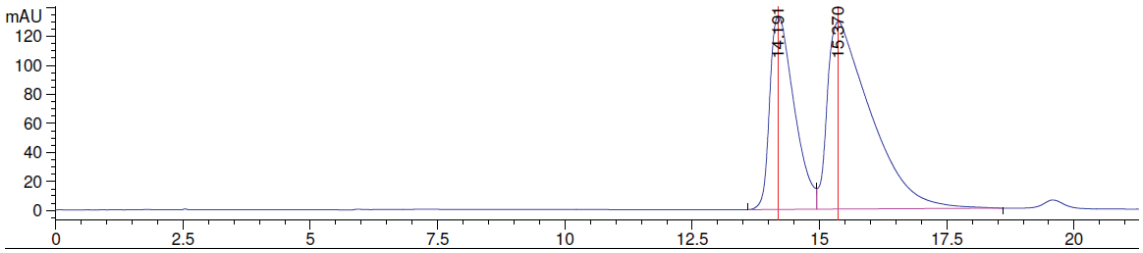
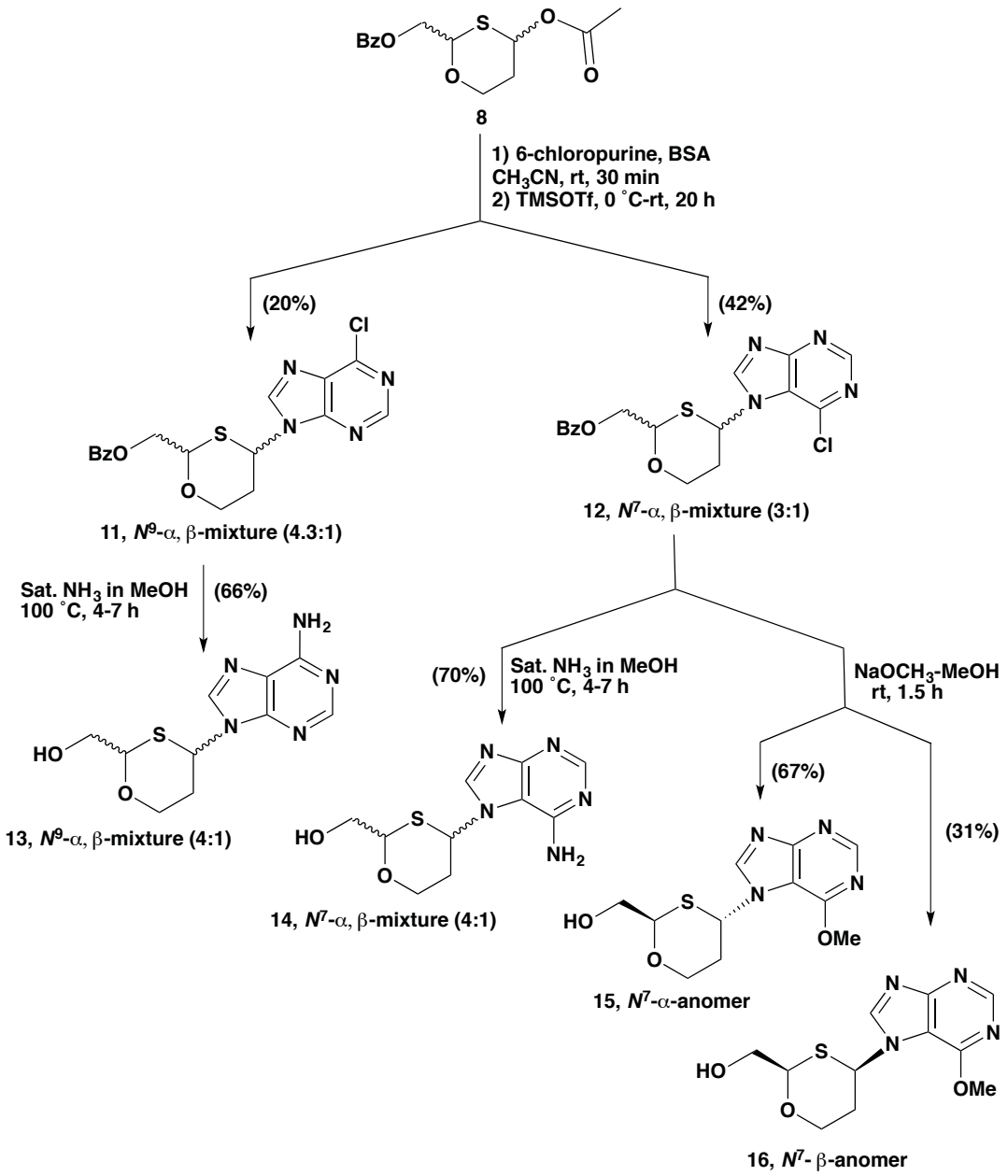


Figure AI.3: RP-HPLC separation of anomeric mixtures of nucleoside **10a**. Conditions: PRP-1 Hamilton column (10 μm, 10 x 250 mm), 8% ACN/H_2O over 25 minutes, 4 ml/min, 25 °C.



Scheme AI.2: Synthesis of 1,3-oxathiane purine nucleosides.

Treatment of N^9-α/β-6-Cl-purine (**11**) and N^7-α/β-6-Cl-purine (**12**) with saturated ammonia in methanol at 100 ^oC in a sealed tube furnished the corresponding adenine derivatives **13** and **14** as a mixture of anomers in good yields. Compound **12** was also treated with a solution of sodium methoxide in methanol to afford N^7-α-OMe-purine (**15**) and N^7-β-OMe-purine (**16**). Anomers **15** and **16** had significantly different polarity and were separated by silica gel column chromatography and resulted approximately in a 2 to 1 ratio.

The relative configuration of the benzoyloxy substituent and the nucleobase was established by NOESY NMR experiments showing the α-anomer as the major product. The NOESY spectra of the *alpha* anomers show a correlation between H2' of the sugar and H8 of purine nucleobase, and between H2' and H6 in pyrimidine nucleosides (**Figure AI.4**). Correlations between H2' and H4' of the sugar ring were observed for the *beta* anomers as expected (**Figure AI.5**). Treatment of the protected 1,3-oxathiane pyrimidine nucleoside analogues **9a-c** with saturated ammonia in methanol furnished the free nucleoside derivatives **10a-c** in high yields.

To determine whether the attachment of 6-chloropurine occurs *via* N^7 or N^9 to the 1,3-oxathiane ring moiety, 2D HMBC NMR and UV experiments were performed.[31] In the case of nucleoside **13**, H2 and H8 of the adenine along with H4' of the sugar moiety showed correlation ($^3J_{CH}$) with C4 of the nucleobase (**Figure AI.6**). This correlation can only be expected for the N^9 regioisomer. In the case of compound **16**, only H2 and H8 of the nucleobase showed correlation ($^3J_{CH}$) with C4 (**Figure AI.7**). Moreover, H4' of the sugar shows a correlation with C5 suggesting attachment *via* N^7.

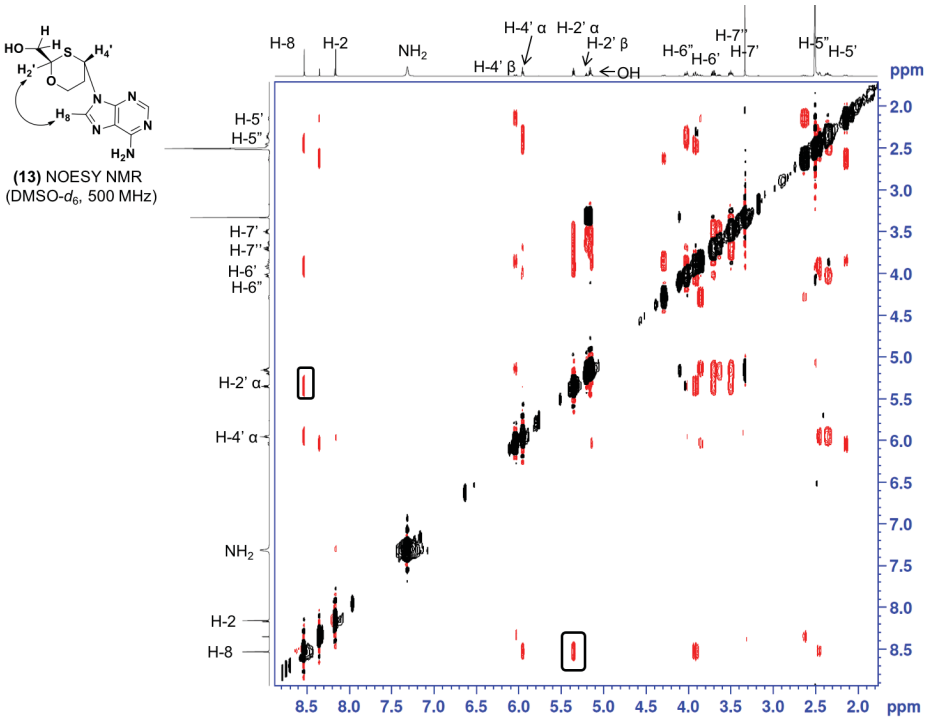


Figure AI.4: NOESY NMR spectrum of 9-(2-hydroxymethyl-1,3-oxathian-4-yl)-adenine (**13**).

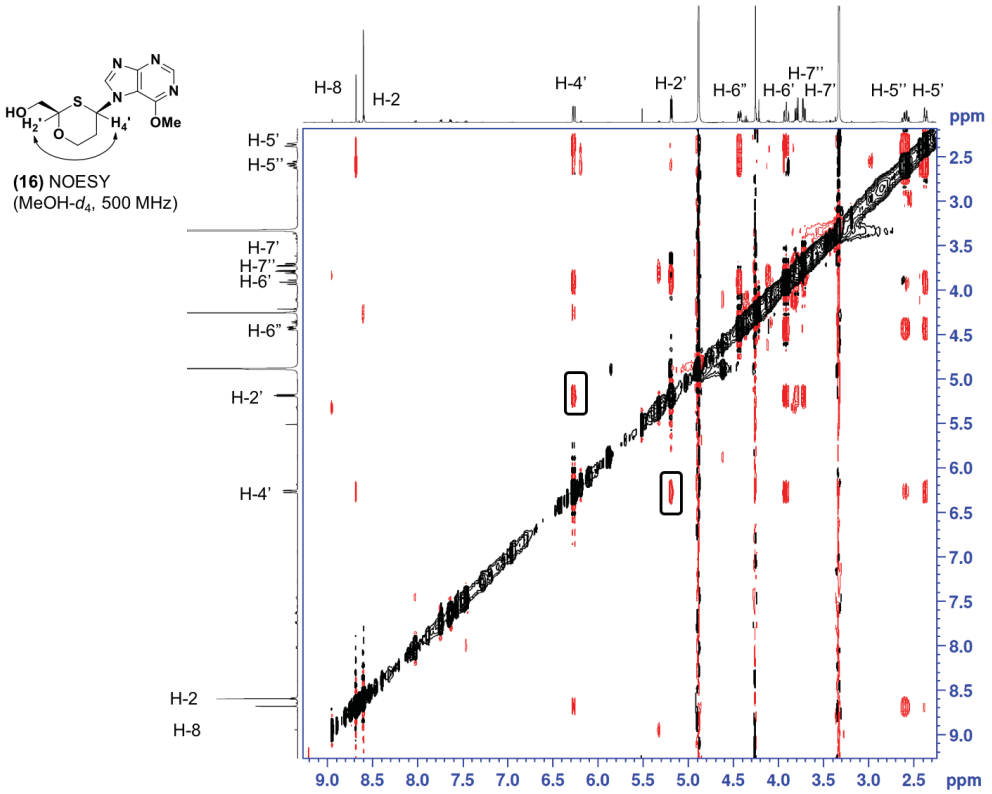


Figure AI.5: NOESY NMR spectrum of β-7-(2-hydroxymethyl-1,3-oxathian-4-yl)-6-methoxypurine (**16**).

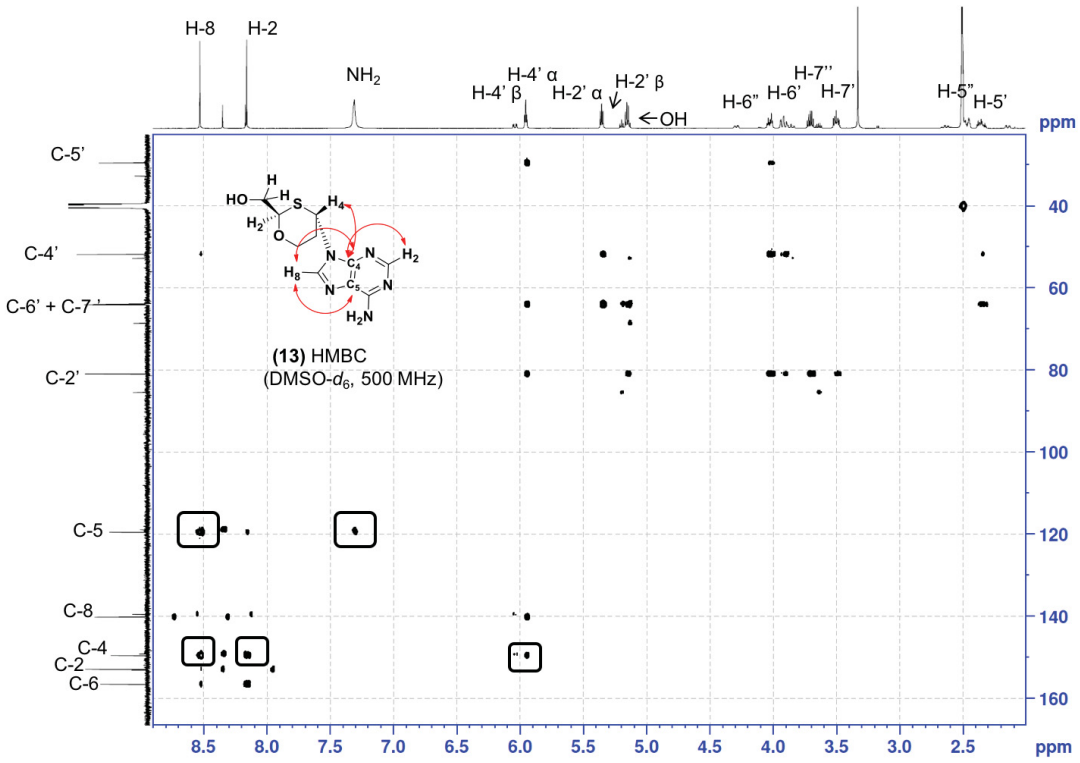


Figure AI.6: HMBC NMR spectrum of 9-(2-hydroxymethyl-1,3-oxathian-4-yl)-adenine (**13**).

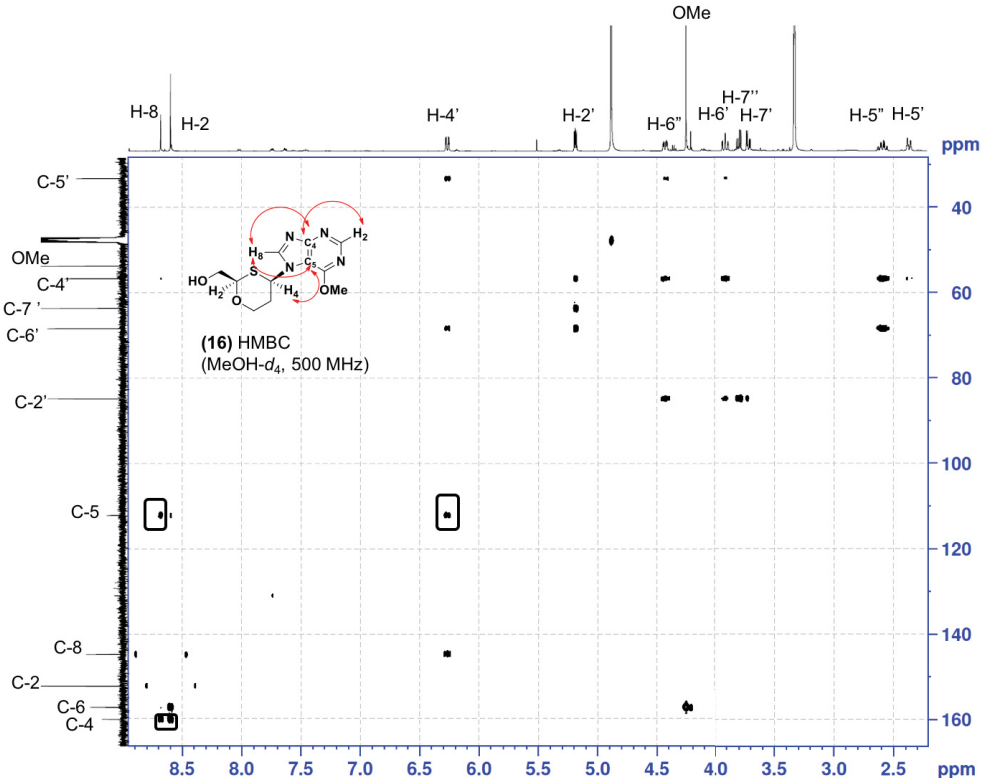


Figure AI.7: HMBC NMR spectrum of β-7-(2-hydroxymethyl-1,3-oxathian-4-yl)-6-methoxypurine (**16**).

The UV spectra of nucleosides follow different patterns depending on the position of attachment of the nucleobase to the sugar moiety. The UV maxima of **13** (261 nm) and **14** (271 nm) were compared with the UV of riboadenosine (260 nm) (**Figure AI.8**), suggesting that nucleoside **13** is attached through N^9 to the sugar moiety, in agreement with HMBC NMR data.

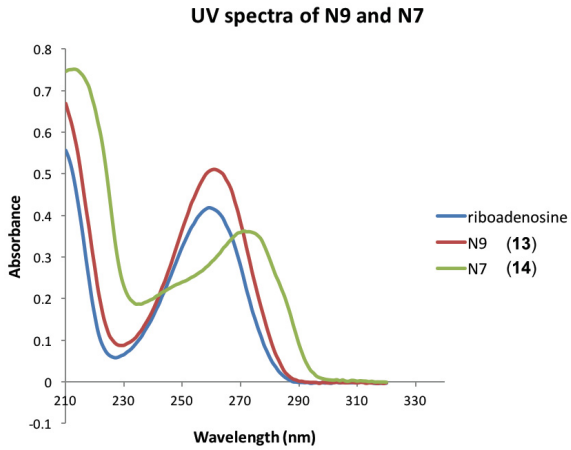


Figure AI.8: UV spectra of nucleosides **13** and **14** in comparison with riboadenosine.

AI.3 Crystal Structure and Conformational Analysis of 1,3-Oxathiane Nucleoside Analogues

The identity, relative stereochemistry, and conformation of N^7-β-OMe-purine (**16**) was unambiguously assigned by X-ray crystallography performed by Dr. Zhijie Chua in the laboratory of Prof. D. Scott Bohle in the Chemistry Department at McGill University. **Figure AI.9** shows the structure of the (S,S) enantiomer of nucleoside **16**. The 1,3-oxathiane moiety crystalizes in a chair conformation where the hydroxymethyl and nucleobase substituents exist in equatorial positions to avoid 1,3-diaxial interactions. In order to study the conformation of N^7-α-OMe-purine (**15**) and N^7-β-OMe-purine (**16**) in solution, we extracted the $^3J_{HH}$ from ^{1}H-NMR spectra for both anomers. The Karplus equation predicts $^3J_{HH}$ in the range 8-12 Hz for protons in a trans-diaxial orientation, while ranges of 2-6 Hz and 0-4 Hz are expected for axial-equatorial and equatorial-equatorial orientations, respectively.

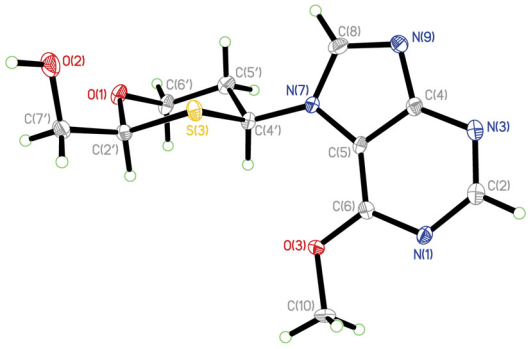


Figure AI.9: Crystal structure of 7-[(2S, 4S)-2-hydroxymethyl-1,3-oxathian-4-yl]-6-methoxypurine (**16**). Only one of the two independent molecules in the asymmetric unit is shown. Figures with the two independent molecules and crystal packing are available in the experimental section.

To study the preferred conformation of the novel 6-OMe-purine analogues, we examined the coupling constants of the H4' anomeric proton (**Figure AI.10**). The ^{1}H-NMR spectra of **16** shows the H4' resonance as a doublet of doublets with different coupling constants ($^3J_{H4'H5''}$ = 12 Hz, $^3J_{H4'H5'}$ = 3.0 Hz) indicating that in solution the conformational equilibrium is shifted towards the conformer observed in the crystal structure (**16A**). On the other hand, in N^7-α-OMe-purine (**15**), the nucleobase and the hydroxymethyl substituents are disposed in a trans relative orientation, and therefore one should adopt an axial position while the other adopts an equatorial

orientation. The anomeric effect should introduce bias towards the **15B** conformation. In the ^{1}H-NMR spectra of N^{7}-α-OMe-purine (**15**), H4' appears as triplet (a doublet of doublets with similar *J* values, $^{3}J_{HH}$ = 3.7 Hz). The absence of a large coupling constant as in the *beta* anomer suggests that the major conformer is indeed **15B**.

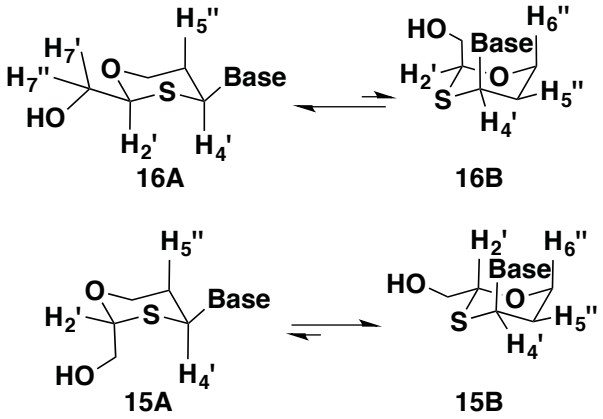


Figure AI.10: The two possible conformations of 1,3-oxathiane nucleosides **15** and **16** in solution.

Table AI.1: ^{1}H-^{1}H Coupling Constant Values for Nucleoside (**15**) and Nucleoside (**16**) at 25 ^{o}C in MeOH-d_4 (500 MHz).

Nucleoside 16	*J* (Hz) (± 0.2) (MeOH-d_4)	*J* (Hz) (± 0.2) (MeOH-d_4)	Nucleoside 15
H4'-H5'	3.0	3.7	H4'-H5'
H4'-H5"	12	3.7	H4'-H5"
H5'-H5"	12.5	15	H5'-H5"
H5"-H6'	12	3	H5"-H6"
H5"-H6"	4	3.5	H5"-H6'
H5'-H6'	2	13	H6'-H6"
H5'-H6"	2.3	4	H6'-H5'

The synthesized compounds were tested for their potency against HIV-1 in an *in vitro* assay in Prof. Mark Wainberg's lab in the Department of Microbiology and Immunology at McGill University. Despite their structural similarity with 3TC, no significant antiviral activity was detected. Interestingly, at the same time as the publication of our work, Brasili and co-workers reported a diastereoselective synthesis of highly related (1,3-dioxan-4-yl)pyrimidine and purine nucleoside analogues. No antiviral activity was observed for these analogues either.[32] The

lack of activity of these six-membered ring nucleosides could be due to inefficient 5'-phosphorylation by kinases and/or their inefficient incorporation by HIV transcriptase. To distinguish between these two scenarios, future work on this class of ring expanded analogues should be centered on the synthesis of 5'-phosphorylated prodrugs to overcome the potentially limiting first phosphorylation step.[33-36]

AI.4 Conclusions

In summary, we developed an efficient route for the synthesis of a set of novel 1,3-oxathiane six-membered nucleoside analogues from readily available starting materials *via* a ring closing pathway. 1,3-oxathiane nucleosides were obtained as an α/β-anomeric mixture with the *alpha* anomers as major products. Glycosylation of the 1,3-oxathiane moiety with 6-chloropurine gave rise to N^7/N^9 glycosylation products. The position of attachment of the nucleobase to the sugar ring was determined by HMBC NMR and UV experiments. The X-ray structure of the N^7-β-OMe-purine derivative (**16**) shows that the 1,3-oxathiane moiety adopts a chair conformation with the hydroxymethyl and nucleobase substituents located in equatorial positions. Despite their structural similarity with the nucleoside analogue 3TC, no significant anti-HIV activity was detected for these compounds.

AI.5 Experimental Section

General Consideration: The reagents used throughout the synthesis were of the highest commercial purity and were used without any further purification. Non-aqueous reactions were performed under anhydrous conditions. All reactions were monitored by TLC and visualized by UV light and/or stained using cerium ammonium molybdate (CAM). Analytical thin-plate chromatography was performed on pre-coated 200 μm layer thick silica gel TLA-R10011B-323 plates (Silicycle). Purifications by column chromatography were performed on silica gel (40-63 μm, 230-400 mesh). High-resolution mass spectra (HRMS) were obtained on a mass spectrometer under electron spray ionization (ESI) conditions.

AI.5.1 2-benzoyloxymethyl-1,3-oxathiane (6):

Starting materials **4** and **5** were dried by co-evaporation three times with benzene. To a cooled solution of 3-mercapto-1-propanol (**5**) (0.5 g, 5.43 mmol) and **4** (0.98 g, 5.97 mmol) in dry CH_2Cl_2 (17 ml), $BF_3.Et_2O$ (0.77 g, 5.43 mmol) was added dropwise. The reaction was stirred for 16 hours at room temperature, then quenched with saturated aqueous $NaHCO_3$. The solution was then filtered over celite, washed with brine and extracted with CH_2Cl_2. The resulting residue was purified by flash chromatography (5% EtOAc/hexane) to afford compound **6** in 65%. R_f (25% EtOAc/hexane): 0.5. ^{1}HNMR ($CDCl_3$, 500 MHz): δ (ppm) 1.74 (m, 1H, H-5), 2.01 (m, 1H, H-5'), 2.85 (dtd, J_{HH} 1.5 Hz, J_{HH} 4.0 Hz, J_{HH} 13.5 Hz, 1H, H-4), 3.09 (td, J_{HH} 3.0 Hz, J_{HH} 13.0 Hz, 1H, H-4'), 3.67 (td, J_{HH} 2.0 Hz, J_{HH} 12.5 Hz, 1H, H-6), 4.23 (m, 1H, H-6'), 4.44 (dd, J_{HH} 4.0 Hz, J_{HH} 11.5 Hz, 1H, H-7), 4.53 (dd, J_{HH} 6.5 Hz, J_{HH} 12.0 Hz, 1H, H-7'), 5.18 (dd, J_{HH} 4.0 Hz, J_{HH} 6.5 Hz , 1H, H-2), 7.46 (t, J_{HH} 8.0 Hz, 2H, Bz), 7.58 (tt, J_{HH} 1.0 Hz, J_{HH} 7.5 Hz, 1H, Bz), 8.09 (m, 2H, Bz). ^{13}C NMR ($CDCl_3$, 125 MHz): δ 25.7 (C-5), 27.6 (C-4), 66.2 (C-7), 69.9 (C-6), 80.7 (C-2), 128.4 (Bz), 129.8 (Bz), 133.1 (Bz), 166.2 (C=O). HRMS (ESI^+) calculated for $C_{12}H_{14}O_3SNa$ $[M+Na]^+$ 261.0561, found 261.05619.

AI.5.2 2-benzoyloxymethyl-1,3-oxathian-3-one (7):

A mixture of compound **6** (0.9 g, 3.38 mmol) and $NaIO_4$ (1.44 g, 6.76 mmol) was dissolved in 10 ml of water and 20 ml of p-dioxane. The reaction was allowed to stir at room temperature for 8 hours. The reaction was filtered over celite to remove solid residues, after which the organic layer was washed with brine and extracted with EtOAc. The product was purified with column chromatography (100% EtOAc) to afford compound **7** in 70% yield as a white solid. R_f (2.4% MeOH /EtOAc): 0.31. major isomer: ^{1}HNMR ($CDCl_3$, 500 MHz) δ (ppm) 2.11 (m, 2H, H-5, H-5'), 2.84 (td, J_{HH} 4.0 Hz, J_{HH} 12.0 Hz, 1H, H-4), 3.68 (m, 2H, H-4, H-6), 4.13 (m, 1H, H-6'), 4.39 (dd, J_{HH} 2.0 Hz, J_{HH} 4.5 Hz, 1H, H-2), 4.81 (dd, J_{HH} 5.0 Hz, J_{HH} 12.5 Hz, 1H, H-7), 4.93 (dd, J_{HH} 2.0 Hz, J_{HH} 12.5 Hz, 1H, H-7), 7.47 (t, J_{HH} 8.0 Hz, 2H, Bz), 7.60 (m, 1H, Bz), 8.08 (m, 2H, Bz). ^{13}C NMR ($CDCl_3$, 125 MHz) δ 24.3 (C-5), 50.7 (C-4), 61.5 (C-7), 69.5 (C-6), 96.0 (C-2), 128.4 (Bz), 129.9 (Bz), 133.3 (Bz), 166.0 (C=O). minor isomer: ^{1}HNMR ($CDCl_3$, 500 MHz) δ 1.61 (m, 1H, H-5), 2.73 (m, 2H, H-4, H-5'), 3.28 (m, 1H, H-4'), 3.80 (m, 1H, H-6), 4.26 (m, 1H, H-6'), 4.60 (m, 2H, H-2, H-7), 4.69 (m, 1H, H-7'), 7.47 (t, J_{HH} 8.0 Hz, 2H, Bz), 7.60 (m, 1H, Bz), 8.08 (m, 2H, Bz).^{13}C NMR ($CDCl_3$, 125 MHz) δ 14.2 (C-5), 44.3 (C-

4), 61.6 (C-7), 69.3 (C-6), 88.2 (C-2), 128.5 (Bz), 129.5 (Bz), 133.5 (Bz), 166.0 (C=O). HRMS (ESI^+) calculated for $C_{12}H_{14}O_4SNa$ $[M+Na]^+$ 277.0510**,** found 277.0507.

AI.5.3 2-benzoyloxymethyl-4-acetoxy-1,3-oxathiane (8):

To a suspension of **7** (0.21 g, 0.84 mmol) in 15 ml Ac_2O, (0.51 g, 1.68 mmol) of $(nBu)_4NOAc$ was added. The reaction was refluxed for 9 hours at 125-130 ^oC. Ac_2O was evaporated and then the compound was neutralized with $NaHCO_3$. Following an aqueous workup, purification with column chromatography (10% EtOAc/hexane) afforded compound **8** in 67% as a pale-yellow gummy liquid. R_f (25% EtOAc/hexane): 0.27. ^{1}HNMR ($CDCl_3$, 500 MHz): major isomer δ (ppm) 2.03 (m, 1H, H-5), 2.16 (s, 3H, CH_3), 2.22 (m, 1H, H-5'), 3.76 (ddd, J_{HH} 3.0 Hz, J_{HH} 8.5 Hz, J_{HH} 12.0 Hz, 1H, H-6), 4.35 (ddd, J_{HH} 3.5 Hz, J_{HH} 6.5 Hz, J_{HH} 12.0 Hz, 1H, H-6'), 4.47 (dd, J_{HH} 4.0 Hz, J_{HH} 12.0 Hz, 1H, H-7), 4.86 (dd, J_{HH} 7.5 Hz, J_{HH} 12.0 Hz, 1H, H-7'), 5.21 (dd, J_{HH} 4.0 Hz, J_{HH} 7.5 Hz 1H, H-2), 6.09 (dd, J_{HH} 3.5 Hz, J_{HH} 8.0 Hz, 1H, H-4), 7.47 (t, J_{HH} 8.0 Hz, 2H, Bz), 7.60 (t, J_{HH} 8.0 Hz, 1H, Bz), 8.09 (d, J_{HH} 8.0 Hz, 2H, Bz). Minor isomer δ 1.85 (m, 1H, H-5), 2.18 (s, 3H, CH_3,) 2.22 (m, 1H, H-5'), 4.13 (m, 2H, H-6, H-6'), 4.45 (dd, J_{HH} 4.0 Hz, J_{HH} 12.0 Hz, 1H, H-7), 4.52 (dd, J_{HH} 6.5 Hz, J_{HH} 11.5 Hz, 1H, H-7'), 5.55 (dd, J_{HH} 4.0 Hz, J_{HH} 6.5 Hz, 1H, H-2), 6.12 (t, J_{HH} 5.0 Hz, 1H, H-4), 7.47 (t, J_{HH} 8.0 Hz, 2H, Bz), 7.60 (t, J_{HH} 8.0 Hz, 1H, Bz), 8.09 (d, J_{HH} 8.0 Hz, 2H, Bz).^{13}C NMR ($CDCl_3$, 125 MHz): δ 21.2 (CH_3), 31.5 (C-5), 63.4 (C-6), 64.9 (C-7), 71.3 (C-4), 78.2 (C-2), 128.4 (Bz), 129.8 (Bz), 133.2 (Bz), 166.0 (C=O), 169.3 (C=O). HRMS (ESI^+) calculated for $C_{14}H_{16}O_5SNa$ $[M+Na]^+$ 319.0616**,** found 319.06188.

AI.5.4 General Procedure for the Glycosylation of Compound 8 with Pyrimidine Bases Followed by Deprotection to Afford Compounds 10a-c:

To a stirred solution of **8** (0.3 g, 1.01 mmol) and the different bases (4.04 mmol) in dry MeCN was added BSA (1.5 ml, 6.07 mmol). The reaction mixture was stirred at room temperature for 30 minutes. This solution was cooled to 0 ^oC and TMSOTf (0.37 ml, 2.02 mmol) was added drop wise. The solution was stirred overnight at room temperature. The resulting residue was purified by column chromatography (0.5-6% MeOH/CH_2Cl_2) to afford the protected nucleosides **9a-c**. The protected nucleosides where deprotected with saturated ammonia in methanol for 16-24 hours at room temperature. Solvent was evaporated and the product was

purified with column chromatography (5-10% MeOH/CH_2Cl_2) to afford the pure 1,3-oxathiane nucleosides **10a-c** as white solids.

AI.5.4.1 1-(2-benzoyloxymethyl-1,3-oxathian-4-yl)-cytosine (9a):

Yield = 54%. R_f(9% MeOH/CH_2Cl_2): 0.33. ^{1}HNMR (MeOH-d_4, 500 MHz): 64% α-isomer δ (ppm) 2.22 (m, 2H, H-5', H-5''), 3.88 (m, 1H, H-6'), 4.13 (dt, J_{HH} 5.0 Hz, J_{HH} 12.5 Hz, 1H, H-6''), 4.50 (m, 1H, H-7'), 4.77 (dd, J_{HH} 7.5 Hz, J_{HH} 12.0 Hz, 1H, H-7''), 5.56 (dd, J_{HH} 4.0 Hz, J_{HH} 7.5 Hz, 1H, H-2'), 6.01 (m, 2H, H-4', H-5), 7.50 (m, 2H, Bz), 7.64 (m, 1H, Bz), 8.07 (td, J_{HH} 1.0 Hz, J_{HH} 7.5 Hz, 2H, Bz), 8.12 (d, J_{HH} 7.5 Hz, 1H, H-6). ^{13}CNMR (MeOH-d_4, 125 MHz): δ 29.3 (C-5'), 53.8 (C-4'), 62.9 (C-6'), 64.7 (C-7'), 76.7 (C-2'), 94.0 (C-5), 128.2 (Bz), 129.3 (Bz), 129.5 (Bz), 133.1 (Bz), 142.7 (C-6), 156.3 (C-2), 165.9 (C-4), 166.0 (C=O, Bz). 36% β-isomer (MeOH-d_4, 500 MHz): δ 1.99 (dq, J_{HH} 2.0 Hz, J_{HH} 13.0 Hz, 1H, H-5'), 2.22 (m, 1H, H-5''), 3.88 (m, 1H, H-6'), 4.37 (ddd, J_{HH} 2.5 Hz, J_{HH} 4.0 Hz, J_{HH} 12.5 Hz, 1H, H-6''), 4.50 (m, 1H, H-7'), 4.58 (dd, J_{HH} 5.5 Hz, J_{HH} 11.5 Hz, 1H, H-7''), 5.43 (dd, J_{HH} 4.5 Hz, J_{HH} 6.0 Hz, 1H, H-2'), 6.01 (m, 1H, H-5), 6.17 (dd, J_{HH} 2.5 Hz, J_{HH} 11.5 Hz, 1H, H-4'), 7.50 (m, 2H, Bz), 7.64 (m, 1H, Bz), 7.72 (d, J_{HH} 7.5 Hz, 1H, H-6), 8.07 (td, J_{HH} 1.0 Hz, J_{HH} 7.5 Hz, 2H, Bz). ^{13}CNMR (MeOH-d_4, 125 MHz): δ 31.8 (C-5'), 54.8 (C-4'), 65.2 (C-7'), 68.7 (C-6'), 81.5 (C-2'), 95.5 (C-5), 128.2 (Bz), 129.3 (Bz), 129.5 (Bz), 133.1 (Bz), 141.8 (C-6), 156.0 (C-2), 165.7 (C-4), 166.0 (C=O). HRMS (ESI^+) calculated for $C_{16}H_{17}N_3O_4SNa$ $[M+Na]^+$ 370.0837**,** found 370.0831.

AI.5.4.2 1-(2-hydroxymethyl-1,3-oxathian-4-yl)-cytosine (10a):

Yield = 86%. R_f(16% MeOH/CH_2Cl_2): 0.27. ^{1}HNMR (MeOH-d_4, 500 MHz): 63% α-isomer δ (ppm) 2.22 (m, 2H, H-5', H-5''), 3.69 (m, 1H, H-7'), 3.80 (m, 1H, H-6'), 3.86 (dd, J_{HH} 6.0 Hz, J_{HH} 12.0 Hz, 1H, H-7''), 4.06 (dt, J_{HH} 5.0 Hz, J_{HH} 12.5 Hz, 1H, H-6''), 5.21 (dd, J_{HH} 4.5 Hz, J_{HH} 6.0 Hz, 1H, H-2'), 5.88 (t, J_{HH} 5.0 Hz, 1H, H-4'), 6.06 (d, J_{HH} 8.0 Hz, 1H, H-5), 8.31 (d, J_{HH} 7.5 Hz, 1H, H-6). ^{13}CNMR (MeOH-d_4, 125 MHz): δ 28.7 (C-5'), 54.9 (C-4'), 62.9 (C-6'), 63.6 (C-7'), 79.8 (C-2'), 94.1 (C-5), 144.4 (C-6), 153.4 (C-2), 163.8 (C-4).

37% β-isomer (MeOH-d_4, 500 MHz): δ (ppm) 1.99 (dq, J_{HH} 2.5 Hz, J_{HH} 13.0 Hz, 1H, H-5'), 2.22 (m, 1H, H-5''), 3.47 (m, 1H, H-7'), 3.69 (m, 1H, H-7''), 3.80 (m, 1H, H-6'), 4.35 (ddd, J_{HH} 2.5 Hz, J_{HH} 4.0 Hz, J_{HH} 12.0 Hz, 1H, H-6''), 5.09 (dd, J_{HH} 4.0 Hz, J_{HH} 5.5 Hz, 1H, H-2'), 6.03 (d, J_{HH} 7.5 Hz, 1H, H-5), 6.10 (dd, J_{HH} 3.0 Hz, J_{HH} 11.5 Hz, 1H, H-4'), 7.81 (d, J_{HH} 7.5 Hz, 1H,

H-6). ^{13}CNMR (MeOH-d_4, 125 MHz): δ 31.8 (C-5'), 54.7 (C-4'), 63.6 (C-7'), 63.7 (C-6'), 84.7 (C-2'), 95.4 (C-5'), 142.9 (C-6), 153.9 (C-2), 164.1 (C-4). HRMS (ESI$^+$) calculated for $C_9H_{13}N_3O_3SNa$ $[M+Na]^+$ 266.0575**,** found 266.0569.

AI.5.4.3 1-(2-benzoyloxymethyl-1,3-oxathian-4-yl)-thymine (9b):

Yield = 68%. R_f(9% MeOH/CH_2Cl_2): 0.64. ^{1}HNMR ($CDCl_3$, 500 MHz): 66 % α-isomer δ (ppm) 2.00 (d, J_{HH} 1.0 Hz, 3H, CH_3), 2.21 (m, 2H, H-5', H-5''), 3.93 (ddd, J_{HH} 3.5 Hz, J_{HH} 7.0 Hz, J_{HH} 12.5 Hz, 1H, H-6'), 4.16 (ddd, J_{HH} 3.5 Hz, J_{HH} 7.0 Hz, J_{HH} 12.5 Hz, 1H, H-6'',), 4.52 (m, 1H, H-7'), 4.86 (dd, J_{HH} 8.0 Hz, J_{HH} 12.0 Hz, 1H, H-7''), 5.46 (dd, J_{HH} 4.5 Hz, J_{HH} 8.0 Hz, 1H, H-2'), 6.00 (dd, J_{HH} 4.5 Hz, J_{HH} 7.5 Hz, 1H, H-4'), 7.48 (m, 2H, Bz), 7.55 (d, J_{HH} 1.0 Hz, 1H, H-6), 7.61 (m, 1H, Bz), 8.09 (m, 2H, Bz), 8.35 (br s, 1H, NH). ^{13}CNMR ($CDCl_3$, 125 MHz): δ 12.8 (CH_3), 30.2 (C-5'), 52.6 (C-4'), 62.9 (C-6'), 64.5 (C-7'), 111.4 (C-5), 128.5 (Bz), 129.4 (Bz), 129.8 (Bz), 133.4 (Bz), 136.4 (C-6), 150.0 (C-2), 163.0 (C-4), 166.0 (C=O). 34% β-isomer ^{1}HNMR ($CDCl_3$, 500 MHz): δ 1.97 (d, J_{HH} 1.0 Hz, 3H, CH_3), 2.04 (q, J_{HH} 2.0 Hz, 1H, H-5'), 2.21 (m, 1H, H-5''), 3.79 (td, J_{HH} 2.0 Hz, J_{HH} 12.0 Hz, 1H, H-6'), 4.41 (ddd, J_{HH} 2.5 Hz, J_{HH} 4.0 Hz, J_{HH} 12.5 Hz, 1H, H-6''), 4.52 (m, 1H, H-7'), 4.63 (m, 1H, H-7''), 5.33 (dd, J_{HH} 4.5 Hz, J_{HH} 6.0 Hz, 1H, H-2'), 6.08 (dd, J_{HH} 3.0 Hz, J_{HH} 12.0 Hz, 1H, H-4'), 7.24 (d, J_{HH} 1.0 Hz, 1H, H-6), 7.48 (m, 2H, Bz), 7.61 (m, 1H, Bz), 8.09 (m, 2H, Bz), 8.35 (br s, 1H, NH). ^{13}CNMR ($CDCl_3$, 125 MHz): δ 12.6 (CH_3), 32.1 (C-5'), 53.7 (C-4'), 65.2 (C-7'), 68.9 (C-6'), 112.3 (C-5), 128.4 (Bz), 129.4 (Bz), 129.9 (Bz), 133.3 (Bz), 135.6 (C-6), 149.9 (C-2), 162.8 (C-4), 165.9 (C=O). HRMS (ESI$^+$) calculated for $C_{17}H_{18}N_2O_5SNa$ $[M+Na]^+$ 385.0834**,** found 385.0831.

AI.5.4.4 1-(2-hydroxymethyl-1,3-oxathian-4-yl)-thymine (10b):

Yield = 89%. R_f(9% MeOH/CH_2Cl_2): 0.17. ^{1}HNMR (DMSO-d_6, 500 MHz): 65 % α-isomer δ (ppm) 1.84 (d, J_{HH} 1.0 Hz, 3H, CH_3), 2.07 (m, 1H, H-5'), 2.20 (m, 1H, H-5''), 3.49 (m, 1H, H-7'), 3.62 (m, 1H, H-7''), 3.86 (m, 1H, H-6'), 3.98 (m, 1H, H-6''), 5.17 (m, 1H, OH), 5.21 (dd, J_{HH} 5.0 Hz, J_{HH} 6.5 Hz, 1H, H-2'), 5.82 (dd, J_{HH} 5.0 Hz, J_{HH} 7.5 Hz, 1H, H-4'), 7.81 (d, J_{HH} 1.5 Hz, 1H, H-6), 11.39 (br s, 1H, NH). ^{13}CNMR (DMSO-d_6, 125 MHz): δ 12.6 (CH_3), 29.4 (C-5'), 52.3 (C-4'), 63.1 (C-6'), 63.8 (C-7'), 80.1 (C-2'), 110.0 (C-5), 137.9 (C-6), 150.8 (C-2), 164.1 (C-4). 35% β-isomer (DMSO-d_6, 500 MHz):δ 1.79 (d, J_{HH} 1.0 Hz, 3H, CH_3), 2.07 (m, 1H, H-5'), 2.20 (m, 1H, H-5''), 3.49 (m, 1H, H-7'), 3.78 (m, 2H, H-6', H-7''), 4.21 (ddd, J_{HH} 2.0 Hz, J_{HH} 3.5

Hz, J_{HH} 11.5 Hz, 1H, H-6''), 5.06 (dd, J_{HH} 5.0 Hz, J_{HH} 6.0 Hz, 1H, H-2'), 5.17 (m, 1H, OH), 5.93 (dd, J_{HH} 3.0 Hz, J_{HH} 12.0 Hz, 1H, H-4'), 7.61 (d, J_{HH} 1.0 Hz, 1H, H-6), 11.39 (br s, 1H, NH). ^{13}CNMR (DMSO-d_6, 125 MHz): δ 12.4 (CH_3), 31.5 (C-5'), 53.7 (C-4'), 63.9 (C-7'), 68.4 (C-6'), 85.2 (C-2'), 110.8 (C-5), 137.2 (C-6), 150.5 (C-2), 163.9 (C-4). HRMS (ESI^+) calculated for $C_{10}H_{14}N_2O_4SNa$ $[M+Na]^+$ 281.0572**,** found 281.0572.

AI.5.4.5 1-(2-benzoyloxymethyl-1,3-oxathian-4-yl)-5-fluorocytosine (9c):

Yield = 72%. R_f(9% MeOH/CH_2Cl_2): 0.31. 67 % α anomer ^{1}HNMR (MeOH-d_4, 500 MHz): δ (ppm) 2.25 (m, 2H, H-5', H-5''), 3.92 (ddd, J_{HH} 4.0 Hz, J_{HH} 7.5 Hz, J_{HH} 12.0 Hz, 1H, H-6'), 4.13 (ddd, J_{HH} 4.0 Hz, J_{HH} 6.0 Hz, J_{HH} 12.5 Hz, 1H, H-6''), 4.51 (m, 1H, H-7'), 4.78 (dd, J_{HH} 7.5 Hz, J_{HH} 12.0 Hz, 1H, H-7''), 5.59 (dd, J_{HH} 4.0 Hz, J_{HH} 7.0 Hz, 1H, H-2'), 5.97 (td, J_{HH} 1.5 Hz, J_{HH} 6.5 Hz, 1H, H-4'), 7.51 (m, 2H, Bz), 7.65 (m, 1H, Bz), 8.07 (m, 2H, Bz), 8.24 (d, J_{HF} 6.5 Hz, 1H, H-6). ^{13}C NMR (MeOH-d_4, 125 MHz): δ 29.0 (C-5'), 54.3 (C-4'), 62.9 (C-6'), 64.7 (C-7'), 76.5 (C-2'), 126.9 (C-6, J_{CF} 32.5 Hz), 128.2 (Bz), 129.3 (Bz), 133.1 (Bz), 138.0 (C-5), 155.0 (C-2), 158.1 (C-4, J_{CF} 17.1 Hz), 166.0 (C=O). 33% β anomer ^{1}HNMR (MeOH-d_4, 500 MHz): δ 2.01 (dq, J_{HH} 2.5 Hz, J_{HH} 13.5 Hz, 1H, H-5'), 2.25 (m, 1H, H-5''), 3.87 (td, J_{HH} 2.0 Hz, J_{HH} 12.0 Hz, 1H, H-6'), 4.37 (ddd, J_{HH} 2.5 Hz, J_{HH} 4.0 Hz, J_{HH} 12.5 Hz 1H, H-6''), 4.51 (m, 1H, H-7'), 4.58 (dd, J_{HH} 5.5 Hz, J_{HH} 11.5 Hz, 1H, H-7''), 5.43 (dd, J_{HH} 4.0 Hz, J_{HH} 6.0 Hz, 1H, H-2'), 6.14 (ddd, J_{HH} 2.0 Hz, J_{HH} 3.0 Hz, J_{HH} 12.0 Hz, 1H, H-4'), 7.51 (m, 2H, Bz), 7.65 (m, 1H, Bz), 7.92 (d, J_{HF} 7.0 Hz, 1H, H-6), 8.07 (m, 2H, Bz). ^{13}C NMR (MeOH-d_4, 125 MHz): δ 31.6 (C-5'), 55.3 (C-4'), 65.2 (C-7'), 68.6 (C-6'), 81.5 (C-2'), 125.9 (C-6, J_{CF} 29.1 Hz), 128.2 (Bz), 129.3 (Bz), 133.1 (Bz), 138.0 (C-5), 154.7 (C-2), 158.1 (C-4, J_{CF} 17.1 Hz), 165.9 (C=O). HRMS (ESI^+) calculated for $C_{16}H_{16}FN_3O_4SNa$ $[M+Na]^+$ 388.0743**,** found 388.0729.

AI.5.4.6 1-(2-hydroxymethyl-1,3-oxathian-4-yl)-5-fluorocytosine (10c):

Yield = 89%. R_f(9% MeOH/CH_2Cl_2): 0.23. 64 % α anomer ^{1}HNMR (MeOH-d_4, 500 MHz): δ (ppm) 2.24 (m, 2H, H-5', H-5''), 3.76 (m, 2H, H-6', H-7'), 3.89 (dd, J_{HH} 6.5 Hz, J_{HH} 12.0 Hz, 1H, H-7''), 4.05 (dt, J_{HH} 4.5 Hz, J_{HH} 12.5 Hz, 1H, H-6''), 5.22 (dd, J_{HH} 4.5 Hz, J_{HH} 6.5 Hz, 1H, H-2'), 5.84 (td, J_{HH} 1.0 Hz, J_{HH} 5.0 Hz, 1H, H-4'), 8.27 (d, J_{HF} 6.5 Hz, 1H, H-6). ^{13}C NMR (MeOH-d_4, 125 MHz): δ 28.9 (C-5'), 54.6 (C-4'), 62.9 (C-6'), 63.6 (C-7'), 79.7 (C-2'), 127.1 (d, J_{CF} 32.7 Hz, C-6), 136.7 (d, J_{CF} 243.4 Hz, C-5), 155.0 (C-2), 158.1 (d, J_{CF} 13.8 Hz, C-4). 36% β

anomer ^{1}HNMR (MeOH-d_4, 500 MHz): δ 1.98 (dq, J_{HH} 2.0 Hz, J_{HH} 13.0 Hz, 1H, H-5'), 2.15 (qd, J_{HH} 4.5 Hz, J_{HH} 12.0 Hz, 1H, H-5''), 3.76 (m, 3H, H-6', H-7', H-7''), 4.34 (ddd, J_{HH} 2.5 Hz, J_{HH} 4.0 Hz, J_{HH} 12.0 Hz, 1H, H-6''), 5.08 (dd, J_{HH} 4.0 Hz, J_{HH} 5.5 Hz, 1H, H-2'), 6.07 (dt, J_{HH} 2.5 Hz, J_{HH} 11.5 Hz, 1H, H-4'), 7.89 (d, J_{HF} 6.5 Hz, 1H, H-6). ^{13}C NMR (MeOH-d_4, 125 MHz): δ 31.8 (C-5'), 55.2 (C-4'), 63.8 (C-7'), 68.5 (C-6'), 84.6 (C-2'), 126.0 (d, J_{CF} 32.5 Hz, C-6), 137.1 (d, J_{CF} 244.7 Hz, C-5), 154.7 (C-2), 157.9 (d, J_{CF} 13.8 Hz, C-4). HRMS (ESI$^+$) calculated for $C_9H_{13}FN_3O_3S$ $[M+H]^+$ 262.0662**,** found 262.0663.

AI.5.5 General procedure for the synthesis of purine 1,3-oxathiane nucleosides:

Similar Vorbrüggen glycosylation conditions as described for the synthesis of nucleosides **9a-c**. After glycosylation and aqueous workup, the residue was purified by column chromatography (0.5-2% MeOH/CH_2Cl_2) to afford 20% of the N^9 regioisomer **11** as a mixture of anomers and 42% of the N^7 regioisomer **12**. 1,3-oxathiane nucleosides **11** and **12** (75 mg, 0.19 mmol) were treated with a saturated solution of ammonia in MeOH (6 ml) and stirred for 4-7 h at 100 ^{o}C in a sealed tube. The reaction was cooled to room temperature, MeOH was evaporated and the residue was purified by column chromatography (5-10% MeOH/CH_2Cl_2) to afford nucleoside **13** (66%) and **14** (70%) as mixtures of anomers as white solids. Nucleoside **12** (40 mg, 0.10 mmol) was treated with sodium methoxide in methanol (3ml) to afford pure α and β nucleosides, **15** (67%) and **16** (31%) respectively, as white solids.

AI.5.5.1 9-(2-benzoyloxymethyl-1,3-oxathian-4-yl)-6-chloropurine (11):

Yield = 20%. R_f(2% MeOH/CH_2Cl_2): 0.48. 81% α-isomer^{1}HNMR (MeOH-d_4, 400 MHz): δ (ppm) 2.57 (m, 1H, H-5'), 2.75 (m, 1H, H-5''), 3.98 (td, J_{HH} 2.0 Hz, J_{HH} 12.8 Hz, 1H, H-6'), 4.14 (dt, J_{HH} 4.0 Hz, J_{HH} 12.8 Hz, 1H, H-6''), 4.52 (m, 1H, H-7'), 4.65 (dd, J_{HH} 6.0 Hz, J_{HH} 11.6 Hz, 1H, H-7''), 5.82 (dd, J_{HH} 4.0 Hz, J_{HH} 6.0 Hz, 1H, H-2'), 6.16 (t, J_{HH} 4.4 Hz, 1H, H-4'), 7.49 (m, 2H, Bz), 7.62 (tt, J_{HH} 1.6 Hz, J_{HH} 8.8 Hz, 1H, Bz), 8.05 (m, 2H, Bz), 8.76 (s, 1H, H-2), 9.02 (s, 1H, H-8). 19% β-isomer δ 2.57 (m, 1H, H-5'), 2.75 (m, 1H, H-5''), 3.98 (td, J_{HH} 2.0 Hz, J_{HH} 12.8 Hz, 1H, H-6'), 4.45 (m, 1H, H-6''), 4.52 (m, 1H, H-7'), 4.60 (dd, J_{HH} 5.6 Hz, J_{HH} 12.0 Hz,1H, H-7''), 5.55 (dd, J_{HH} 4.0 Hz, J_{HH} 5.6 Hz, 1H, H-2'), 6.31 (dd, J_{HH} 3.2 Hz, J_{HH} 12.0 Hz, 1H, H-4'), 7.49 (m, 2H, Bz), 7.62 (tt, J_{HH} 1.6 Hz, J_{HH} 8.8 Hz, 1H, Bz), 8.05 (m, 2H, Bz), 8.78 (s, 1H, H-2), 8.80 (s, 1H, H-8).

^{13}C NMR (MeOH-d_4, 125 MHz): δ 28.6 (C-5'), 53.9 (C-4'), 63.8 (C-6'), 65.2 (C-7'), 77.5 (C-2'), 128.2 (Bz), 129.3 (Bz), 129.5 (Bz), 133.2 (Bz), 146.2 (C-5), 150.1 (C-8), 151.4 (C-6), 151.6 (C-4), 151.8 (C-2), 166.0 (C=O). HRMS (ESI$^+$) calculated for $C_{17}H_{15}ClN_4O_3SNa$ $[M+Na]^+$ 413.0451**,** found 413.0458.

AI.5.5.2 7-(2-benzoyloxymethyl-1,3-oxathian-4-yl)-6-chloropurine (12):

Yield = 42%. R_f(2% MeOH/CH_2Cl_2): 0.30. 75% α-isomer ^{1}HNMR (MeOH-d_4, 400 MHz): δ (ppm) 2.47 (m, 1H, H-5'), 2.70 (m, 1H, H-5''), 3.99 (qd, J_{HH} 2.4 Hz, J_{HH} 13.2 Hz, 1H, H-6'), 4.19 (dt, J_{HH} 4.0 Hz, J_{HH} 13.2 Hz, 1H, H-6''), 4.52 (dd, J_{HH} 3.6 Hz, J_{HH} 12.0 Hz, 1H, H-7'), 4.65 (dd, J_{HH} 6.0 Hz, J_{HH} 12.0 Hz, 1H, H-7''), 5.67 (dd, J_{HH} 3.6 Hz, J_{HH} 6.0 Hz, 1H, H-2'), 6.50 (t, J_{HH} 3.6 Hz, 1H, H-4'), 7.48 (m, 2H, Bz), 7.62 (m, 1H, Bz), 8.04 (m, 2H, Bz), 8.81 (s, 1H, H-2), 9.24 (s, 1H, H-8). ^{13}C NMR (MeOH-d_4, 125 MHz): δ 29.8 (C-5'), 54.9 (C-4'), 63.5 (C-6'), 65.0 (C-7'), 77.4 (C-2'), 122.2 (C-8), 128.2 (Bz), 129.3 (Bz), 129.4 (Bz), 133.2 (Bz), 143.6 (C-6), 148.7 (C-4), 151.9 (C-2), 161.7 (C-5), 166.0 (C=O). β-isomer ^{1}HNMR (MeOH-d_4, 400 MHz): δ 2.47 (m, 1H, H-5'), 2.70 (m, 1H, H-5''), 3.99 (qd, J_{HH} 2.4 Hz, J_{HH} 13.2 Hz, 1H, H-6'), 4.47 (ddd, J_{HH} 2.4 Hz, J_{HH} 4.0 Hz, J_{HH} 12.4 Hz, 1H, H-6''), 4.54 (m, 1H, H-7'), 4.60 (dd, J_{HH} 5.6 Hz, J_{HH} 12.0 Hz, 1H, H-7''), 5.57 (dd, J_{HH} 4.4 Hz, J_{HH} 5.6 Hz, 1H, H-2'), 6.52 (d, J_{HH} 2.8 Hz, 1H, H-4'), 7.48 (m, 2H, Bz), 7.62 (m, 1H, Bz), 8.04 (m, 2H, Bz), 8.82 (s, 1H, H-2), 9.01 (s, 1H, H-8). ^{13}C NMR (MeOH-d_4, 125 MHz): δ 32.5 (C-5'), 56.3 (C-4'), 65.0 (C-7'), 68.5 (C-6'), 81.7 (C-2'), 121.6 (C-8), 128.2 (Bz), 129.3 (Bz), 129.4 (Bz), 133.2 (Bz), 143.4 (C-6), 148.1 (C-4), 152.1 (C-2), 161.1 (C-5), 165.9 (C=O). HRMS (ESI$^+$) calculated for $C_{17}H_{15}ClN_4O_3SNa$ $[M+Na]^+$ 413.0451**,** found 413.0454.

AI.5.5.3 9-(2-hydroxymethyl-1,3-oxathian-4-yl)-adenine (13):

Yield = 66%. R_f(9% MeOH/CH_2Cl_2): 0.14. UV λ_{max} (H_2O) 261 nm (8529 M^{-1} cm^{-1}). 80 % α-isomer ^{1}HNMR (DMSO-d_6, 500 MHz): δ (ppm) 2.35 (m, 1H, H-5'), 2.47 (m, 1H, H-5''), 3.49 (m, 1H, H-7'), 3.70 (m, 1H, H-7''), 3.91 (m, 1H, H-6'), 4.02 (dt, J_{HH} 4.0 Hz, J_{HH} 12.5 Hz, 1H, H-6''), 5.15 (m, 1H, OH), 5.35 (t, J_{HH} 5.5 Hz, 1H, H-2'), 5.95 (t, J_{HH} 4.5 Hz, 1H, H-4'), 7.31 (br s, 2H, NH_2), 8.16 (s, 1H, H-2), 8.53 (s, 1H, H-8). ^{13}C NMR (DMSO-d_6, 125 MHz): δ 29.5 (C-5'), 51.8 (C-4'), 63.9 (C-7'), 64.0 (C-6'), 80.9 (C-2'), 119.5 (C-5), 140.2 (C-8), 149.6 (C-4), 152. 9 (C-2), 156.6 (C-6). 20% β-isomer δ 2.14 (m, 1H, H-5'), 2.63 (m, 1H, H-5''), 3.49 (m, 1H,

H-7'), 3.64 (m, 1H, H-7''), 3.86 (m, 1H, H-6'), 4.29 (ddd, J_{HH} 2.0 Hz, J_{HH} 3.5 Hz, J_{HH} 12.0 Hz, 1H, H-6''), 5.15 (m, 1H, OH), 5.20 (t, J_{HH} 6.0 Hz, 1H, H-2'), 6.04 (dd, J_{HH} 2.5 Hz, J_{HH} 12.0 Hz, 1H, H-4'), 7.31 (br s, 2H, NH_2), 8.17 (s, 1H, H-2), 8.35 (s, 1H, H-8). ^{13}C NMR (DMSO-d_6, 125 MHz): δ 32.7 (C-5'), 52.7 (C-4'), 63.8 (C-7'), 68.6 (C-6'), 118.9 (C-5), 139.5 (C-8), 149.6 (C-4), 153.1 (C-2), 156.6 (C-6). HRMS (ESI^+) calculated for $C_{10}H_{13}N_5O_2SNa$ $[M+Na]^+$ 290.0682**,** found 290.06836.

AI.5.5.4 7-(2-hydroxymethyl-1,3-oxathian-4-yl)-adenine (14):

Yield = 70%. R_f(16% MeOH/CH_2Cl_2): 0.14.UV λ_{max} (H_2O) 271 nm (4847 M^{-1} cm^{-1}).80% α anomer ^{1}HNMR (MeOH-d_4, 500 MHz): δ (ppm) 2.33 (m, 1H, H-5'), 2.57 (m, 1H, H-5''), 3.78 (m, 2H, H-6', H-7'), 3.89 (m, 1H, H-7''), 4.13 (dt, J_{HH} 4.0 Hz, J_{HH} 13.0 Hz, 1H, H-6''), 5.28 (dd, J_{HH} 4.0 Hz, J_{HH} 5.5 Hz, 1H, H-2'), 6.26 (t, J_{HH} 4.5 Hz, 1H, H-4'), 8.29 (s, 1H, H-2), 8.79 (s, 1H, H-8). ^{13}C NMR (MeOH-d_4, 125 MHz): δ 30.7 (C-5'), 54.6 (C-4'), 62.9 (C-6'), 63.9 (C-7'), 80.4 (C-2'), 111.0 (C-5), 144.6 (C-8), 152.0 (C-6), 152.5 (C-2), 159.7 (C-4). 20% β anomer ^{1}HNMR (MeOH-d_4, 500 MHz): δ 2.33 (m, 1H, H-5'), 2.57 (m, 1H, H-5''), 3.78 (m, 2H, H-6', H-7'), 3.89 (m, 1H, H-7''), 4.45 (ddd, J_{HH} 2.5 Hz, J_{HH} 4.5 Hz, J_{HH} 12.5 Hz, 1H, H-6''), 5.27 (m, 1H, H-2'), 6.18 (dd, J_{HH} 3.5 Hz, J_{HH} 12.0 Hz, 1H, H-4'), 8.31 (s, 1H, H-2), 8.51 (s, 1H, H-8). ^{13}C NMR (MeOH-d_4, 125 MHz): δ 32.8 (C-5'), 56.3 (C-4'), 63.5 (C-7'), 68.3 (C-6'), 84.8 (C-2'), 111.0 (C-5), 144.6 (C-8), 151.7 (C-6), 152.6 (C-2), 159.5 (C-4). HRMS (ESI^+) calculated for $C_{10}H_{13}N_5O_2SNa$ $[M+Na]^+$ 290.0682**,** found 290.06762.

AI.5.5.5 α-7-(2-hydroxymethyl-1,3-oxathian-4-yl)-6-methoxypurine (15):

Yield = 67%. R_f(9% MeOH/CH_2Cl_2): 0.27. ^{1}HNMR (MeOH-d_4, 500 MHz): δ (ppm) 2.41 (dq, J_{HH} 3.0 Hz, J_{HH} 15.0 Hz, 1H, H-5'), 2.61 (m, 1H, H-5''), 3.70 (dd, J_{HH} 4.5 Hz, J_{HH} 12.0 Hz, 1H, H-7'), 3.83 (m, 2H, H-6', H-7''), 4.11 (dt, J_{HH} 4.0 Hz, J_{HH} 13.0 Hz, 1H, H-6''), 4.22 (s, 3H, OCH_3), 5.32 (dd, J_{HH} 4.5 Hz, J_{HH} 5.0 Hz, 1H, H-2'), 6.19 (t, J_{HH} 4.0 Hz, 1H, H-4'), 8.59 (s, 1H, H-2), 8.95 (s, 1H, H-8). ^{13}C NMR (MeOH-d_4, 75 MHz): δ 30.0 (C-5'), 53.7 (C-OMe), 55.5 (C-4'), 63.4 (C-6'), 63.8 (C-7'), 80.3 (C-2'), 112.7 (C-5), 145.2 (C-8), 152.0 (C-2), 157.5 (C-4), 161.0 (C-6). HRMS (ESI^+) calculated for $C_{11}H_{14}N_4O_3SNa$ $[M+Na]^+$ 305.0684**,** found 305.06725.

AI.5.5.6 β-7-(2-hydroxymethyl-1,3-oxathian-4-yl)-6-methoxypurine (16):

Yield = 31%. R_f(9% MeOH/CH_2Cl_2): 0.33. ^{1}HNMR (MeOH-d_4, 500 MHz): δ (ppm) 2.36 (dq, J_{HH} 2.0 Hz, J_{HH} 13.0 Hz, 1H, H-5'), 2.59 (qd, J_{HH} 4.0 Hz, J_{HH} 12.0 Hz, 1H, H-5''), 3.72 (dd, J_{HH} 4.0 Hz, J_{HH} 12.0 Hz, 1H, H-7'), 3.80 (dd, J_{HH} 6.0 Hz, J_{HH} 12.0 Hz, 1H, H-7''), 3.91 (td, J_{HH} 2.0 Hz, J_{HH} 12.0 Hz, 1H, H-6'), 4.25 (s, 3H, OCH_3), 4.43 (ddd, J_{HH} 2.5 Hz, J_{HH} 4.0 Hz, J_{HH} 12.0 Hz, 1H, H-6''), 5.18 (dd, J_{HH} 4.0 Hz, J_{HH} 5.5 Hz, 1H, H-2'), 6.26 (dd, J_{HH} 3.0 Hz, J_{HH} 12.0 Hz, 1H, H-4'), 8.59 (s, 1H, H-2), 8.68 (s, 1H, H-8). ^{13}C NMR (MeOH-d_4, 125 MHz): δ 33.2 (C-5'), 53.7 (C-OMe), 56.7 (C-4'), 63.6 (C-7'), 68.4 (C-6'), 84.8 (C-2'), 112.1 (C-5), 144.7 (C-8), 152.0 (C-2), 157.4 (C-6), 160.1 (C-4). HRMS (ESI^+) calculated for $C_{11}H_{14}N_4O_3SNa$ $[M+Na]^+$ 305.0684**,** found 305.06703.9

AI.5.6 X-ray Crystallography:

Crystals are mounted with Mitegen mounts using Paratone-N from Hampton Research and single-crystal X-ray diffraction experiments are carried out with a BRUKER APEX-II CCD diffractometer by using graphite-monochromated $Mo_{K\alpha}$ radiation (λ = 0.71073 Å) and KRYOFLEX for low temperature experiments. SAINT[37] is used for integration of the intensity reflections and scaling and SADABS[38] for absorption correction. Patterson maps are used to generate the initial solutions. Non-hydrogen atoms are located by difference Fourier maps and final solution refinements are solved by full-matrix least-squares method on F^2 of all data, by using SHELXTL software.[37] The hydrogen atoms are placed in calculated positions. The crystal structure is solved in the polar non-centrosymmetric space group $Pna2_1$ and contains unequal fractions of the two enantiomers of **16** due to racemic twinning. TWIN[37] and BASF[37] parameters are included for further refinement and indicates the (S, S) enantiomer to be the major fraction (75%) with the (R, R) enantiomer as the minor fraction (25%). The Flack parameter,[39,40] taken from BASF coefficient in this case, is 0.25(11). Crystals of nucleoside **16** are obtained from slow evaporation of CD_3OD at room temperature. The cif for the X-ray diffraction data is available at CCDC with the CCDC code number (1039733).

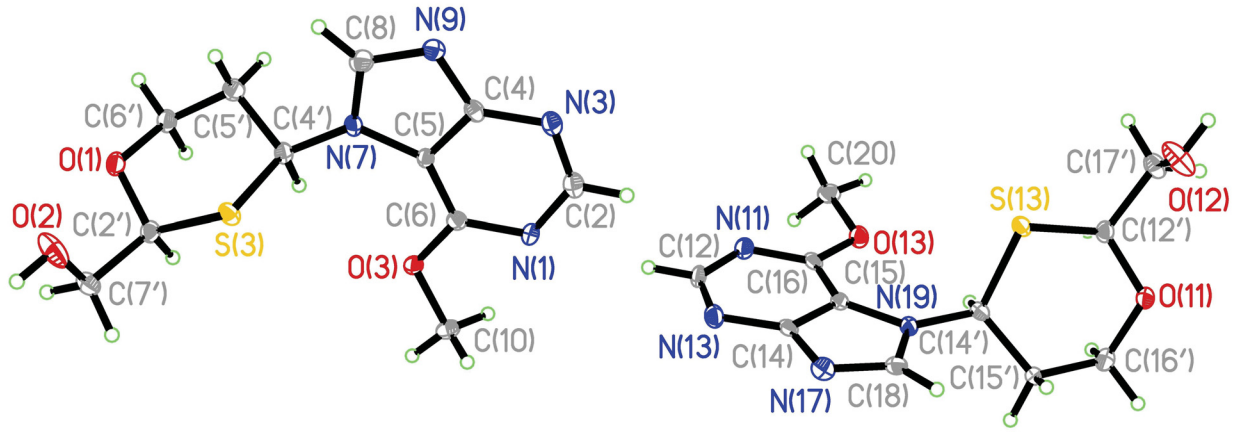


Figure AI.11: Crystal structure of the two independent molecules of 7-[(2S, 4S)-2-hydroxymethyl-1,3-oxathian-4-yl]-6-methoxypurine (nucleoside **16**) in asymmetric unit.

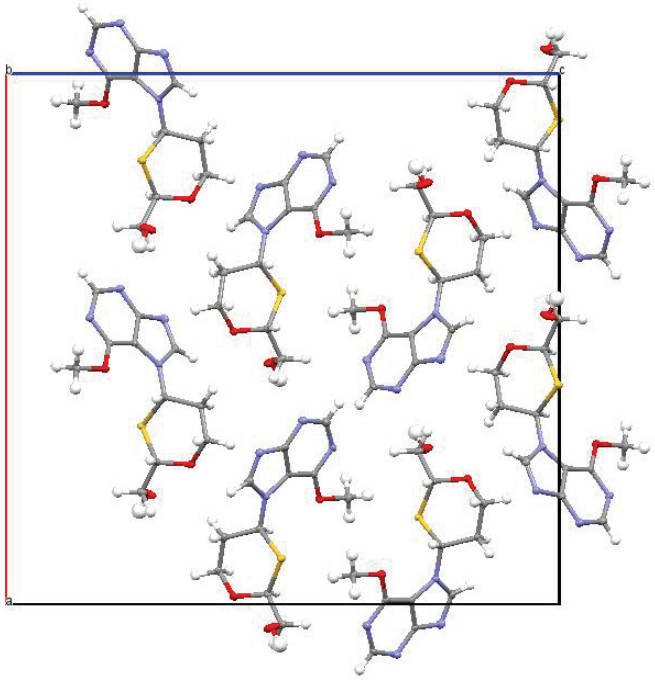


Figure AI.12: Crystal packing of nucleoside **16**.

AI.6 References

1. Parker, W. B., Enzymology of Purine and Pyrimidine Antimetabolites Used in the Treatment of Cancer. *Chem. Rev.* **2009**, *109* (7), 2880-2893.
2. De Clercq, E., A 40-year journey in search of selective antiviral chemotherapy. *Annu. Rev. Pharmacol. Toxicol.* **2011**, *51*, 1-24.
3. Mitsuya, H.; Weinhold, K. J.; Furman, P. A.; Stclair, M. H.; Lehrman, S. N.; Gallo, R. C.; Bolognesi, D.; Barry, D. W.; Broder, S., 3'-Azido-3'-Deoxythymidine (Bw A509u) -

an Antiviral Agent That Inhibits the Infectivity and Cytopathic Effect of Human Lymphotropic-T Virus Type-Iii Lymphadenopathy-Associated Virus Invitro. *Proc. Natl. Acad. Sci. USA* **1985**, *82* (20), 7096-7100.

4. Mitsuya, H.; Broder, S., Inhibition of the Invitro Infectivity and Cytopathic Effect of Human Lymphotrophic-T Virus "Type-Iii/Lymphadenopathy-Associated Virus (Htlv-Iii/Lav) by 2',3'-Dideoxynucleosides. *Proc. Natl. Acad. Sci. USA* **1986**, *83* (6), 1911-1915.
5. Lin, T. S.; Schinazi, R. F.; Prusoff, W. H., Potent and Selective Invitro Activity of 3'-Deoxythymidin-2'-Ene (3'-Deoxy-2',3'-Didehydrothymidine) against Human-Immunodeficiency-Virus. *Biochem. Pharmacol.* **1987**, *36* (17), 2713-2718.
6. Daluge, S. M.; Good, S. S.; Faletto, M. B.; Miller, W. H.; StClair, M. H.; Boone, L. R.; Tisdale, M.; Parry, N. R.; Reardon, J. E.; Dornsife, R. E.; Averett, D. R.; Krenitsky, T. A., 1592U89, a novel carbocyclic nucleoside analog with potent, selective anti-human immunodeficiency virus activity. *Antimicrob. Agents Chemother.* **1997**, *41* (5), 1082-1093.
7. Marquez, V. E.; Siddiqui, M. A.; Ezzitouni, A.; Russ, P.; Wang, J. Y.; Wagner, R. W.; Matteucci, M. D., Nucleosides with a twist. Can fixed forms of sugar ring pucker influence biological activity in nucleosides and oligonucleotides? *J. Med. Chem.* **1996**, *39* (19), 3739-3747.
8. Soudeyns, H.; Yao, X. I.; Gao, Q.; Belleau, B.; Kraus, J. L.; Nguyen-Ba, N.; Spira, B.; Wainberg, M. A., Anti-human immunodeficiency virus type 1 activity and in vitro toxicity of 2'-deoxy-3'-thiacytidine (BCH-189), a novel heterocyclic nucleoside analog. *Antimicrob. Agents Chemother.* **1991**, *35* (7), 1386-1390.
9. Schinazi, R. F.; McMillan, A.; Cannon, D.; Mathis, R.; Lloyd, R. M.; Peck, A.; Sommadossi, J. P.; St Clair, M.; Wilson, J.; Furman, P. A., Selective inhibition of human immunodeficiency viruses by racemates and enantiomers of cis-5-fluoro-1-[2-(hydroxymethyl)-1,3-oxathiolan-5-yl]cytosine. *Antimicrob. Agents Chemother.* **1992**, *36* (11), 2423-2431.
10. Larder, B. A.; Darby, G.; Richman, D. D., Hiv with Reduced Sensitivity to Zidovudine (Azt) Isolated during Prolonged Therapy. *Science* **1989**, *243* (4899), 1731-1734.
11. Arts, E. J.; Wainberg, M. A., Mechanisms of nucleoside analog antiviral activity and resistance during human immunodeficiency virus reverse transcription. *Antimicrob. Agents Chemother.* **1996**, *40* (3), 527-540.
12. Leumann, C. J., DNA analogues: From supramolecular principles to biological properties. *Bioorg. Med. Chem.* **2002,** *10* (4), 841-854.
13. Herdewijn, P., Nucleic Acids with a Six-Membered 'Carbohydrate' Mimic in the Backbone. *Chem. Biodiversity* **2010**, *7* (1), 1-59.
14. Barral, K.; Courcambeck, J.; Pepe, G.; Balzarini, J.; Neyts, J.; De Clercq, E.; Camplo, M., Synthesis and antiviral evaluation of cis-substituted cyclohexenyl and cyclohexanyl-nucleosides. *J. Med. Chem.* **2005**, *48* (2), 450-456.
15. Hronowski, L. J. J.; Szarek, W. A., Synthesis and Biological Evaluation of Novel Pyrimidine Nucleoside Analogs of 1,4-Oxathiane, 1,4-Dithiane, and 1,4-Dioxane. *J. Med. Chem.* **1982**, *25* (5), 522-526.
16. Prisbe, E. J., Synthesis and Antiviral Evaluation of 1,4-Dioxane Nucleoside Analogs Related to Nucleoside Dialdehydes. *J. Med. Chem.* **1986**, *29* (12), 2445-2450.

17. Maeda, M.; Kajimoto, N.; Yamaizumi, Z.; Okamoto, Y.; Nagahara, K.; Takayanagi, H., Synthesis of a novel dioxane nucleoside having two bases, 2(R)-(5-fluorouracil-1-yl)-5(R)-hydroxymethyl-3(R)-(uracil-1-yl)-1,4-dioxane and its 2(S)-isomer, from uridine. *Tetrahedron Lett.* **1997**, *38* (39), 6841-6844.
18. Vanaerschot, A.; Janssen, G.; Herdewijn, P., Nucleoside Analogs with a 1,4-Dioxane, 1,4-Oxathiane or 1,4-Oxazine Ring Structure as the Carbohydrate Fragment. *Bull. Soc. Chim. Belg.* **1990**, *99* (10), 769-777.
19. Van Aerschot, A.; Balzarini, J.; De Clercq, E.; Herdewijn, P., Synthesis and Antiviral Activity of 1.4-Dioxane, 1.4-Oxathiane and 1,4-Morpholine Nucleoside Analogues. *Nucleosides and Nucleotides* **1991**, *10* (1-3), 591-592.
20. Capaldi, D. C.; Eleuteri, A.; Chen, Q.; Schinazi, R. F., Synthesis of six-membered nucleoside analogs .1. Pyrimidine nucleosides based on the 1,3-dioxane ring system. *Nucleosides and Nucleotides* **1997**, *16* (4), 403-416.
21. Watts, J. K.; Sadalapure, K.; Choubdar, N.; Pinto, B. M.; Damha, M. J., Synthesis and conformational analysis of 2 '-fluoro-5-methyl-4 '-thioarabinouridine (4 ' S-FMAU). *J. Org. Chem.* **2006**, *71* (3), 921-925.
22. Sabatino, D.; Damha, M. J., Oxepane nucleic acids: Synthesis, characterization, and properties of oligonucleotides bearing a seven-membered carbohydrate ring. *J. Am. Chem. Soc.* **2007**, *129* (26), 8259-8270.
23. Watts, J. K.; Johnston, B. D.; Jayakanthan, K.; Wahba, A. S.; Pinto, B. M.; Damha, M. J., Synthesis and biophysical characterization of oligonucleotides containing a 4 '-selenonucleotide. *J. Am. Chem. Soc.* **2008**, *130* (27), 8578-+.
24. Habibian, M.; Martinez-Montero, S.; Portella, G.; Chua, Z. J.; Bohle, D. S.; Orozco, M.; Damha, M. J., Seven-Membered Ring Nucleoside Analogues: Stereoselective Synthesis and Studies on Their Conformational Properties. *Org. Lett.* **2015**, *17* (21), 5416-5419.
25. Du, J. F.; Watanabe, K. A., Facile preparation of alpha-acyloxyacetaldehyde, a versatile intermediate in the synthesis of antiviral nucleosides. *Synth. Commun.* **2004**, *34* (11), 1925-1930.
26. Kumar, V.; Dev, S., Titanium Tetrachloride, an Efficient and Convenient Reagent for Thioacetalization. *Tetrahedron Lett.* **1983**, *24* (12), 1289-1292.
27. Karimi, B.; Seradj, H., Zirconium tetrachloride (ZrC4) as an efficient and chemoselective catalyst for conversion of carbonyl compounds to 1,3-oxathiolanes. *Synlett* **2000**, (6), 805-806.
28. Yadav, J. S.; Reddy, B. V. S.; Pandey, S. K., LiBF4 catalyzed chemoselective conversion of aldehydes to 1,3-oxathiolanes and 1,3-dithianes. *Synlett* **2001**, (2), 238-239.
29. Belleau, B.; Brasili, L.; Chan, L.; Dimarco, M. P.; Zacharie, B.; Nguyenba, N.; Jenkinson, H. J.; Coates, J. A. V.; Cameron, J. M., A Novel Class of 1,3-Oxathiolane Nucleoside Analogs Having Potent Anti-Hiv Activity. *Bioorg. Med. Chem. Lett.* **1993**, *3* (8), 1723-1728.
30. Vorbruggen, H.; Krolikiewicz, K.; Bennua, B., Nucleoside Syntheses .22. Nucleoside Synthesis with Trimethylsilyl Triflate and Perchlorate as Catalysts. *Chem. Ber.* **1981**, *114* (4), 1234-1255.
31. Martinez-Montero, S.; Fernandez, S.; Sanghvi, Y. S.; Theodorakis, E. A.; Detorio, M. A.; McBrayer, T. R.; Whitaker, T.; Schinazi, R. F.; Gotor, V.; Ferrero, M., Synthesis,

evaluation of anti-HIV-1 and anti-HCV activity of novel 2',3'-dideoxy-2',2'-difluoro-4'-azanucleosides. *Bioorg. Med. Chem.* **2012**, *20* (23), 6885-93.
32. Battisti, U. M.; Sorbi, C.; Quotadamo, A.; Franchini, S.; Tait, A.; Schols, D.; Jeong, L. S.; Lee, S. K.; Song, J.; Brasili, L., Diastereoselective Synthesis of (1,3-Dioxan-4-yl)pyrimidine and Purin Nucleoside Analogues. *Eur. J. Org. Chem.* **2015**, *2015* (6), 1235-1245.
33. Bobeck, D. R.; Schinazi, R. F.; Coats, S. J., Advances in nucleoside monophosphate prodrugs as anti-HCV agents. *Antiviral therapy* **2010**, *15* (7), 935-50.
34. McGuigan, C.; Bourdin, C.; Derudas, M.; Hamon, N.; Hinsinger, K.; Kandil, S.; Madela, K.; Meneghesso, S.; Pertusati, F.; Serpi, M.; Slusarczyk, M.; Chamberlain, S.; Kolykhalov, A.; Vernachio, J.; Vanpouille, C.; Introini, A.; Margolis, L.; Balzarini, J., Design, synthesis and biological evaluation of phosphorodiamidate prodrugs of antiviral and anticancer nucleosides. *Eur. J. Med. Chem.* **2013**, *70*, 326-40.
35. Velanguparackel, W.; Hamon, N.; Balzarini, J.; McGuigan, C.; Westwell, A. D., Synthesis, anti-HIV and cytostatic evaluation of 3 '-deoxy-3 '-fluorothymidine (FLT) pro-nucleotides. *Bioorg. Med Chem. Lett.* **2014**, *24* (10), 2240-2243.
36. Slusarczyk, M.; Lopez, M. H.; Balzarini, J.; Mason, M.; Jiang, W. G.; Blagden, S.; Thompson, E.; Ghazaly, E.; McGuigan, C., Application of ProTide Technology to Gemcitabine: A Successful Approach to Overcome the Key Cancer Resistance Mechanisms Leads to a New Agent (NUC-1031) in Clinical Development. *J. Med. Chem.* **2014**, *57* (4), 1531-1542.
37. Sheldrick, G. M., A short history of SHELX. *Acta crystallographica. Section A, Foundations of crystallography* **2008**, *64* (Pt 1), 112-22.
38. Sheldrick, G. M., SADABS, TWINABS. University of Gottingen, Germany **1996**.
39. Flack, H. D., On Enantiomorph-Polarity Estimation. *Acta Crystallographica Section A* **1983**, *39*, 876-881.
40. Bernardinelli, G.; Flack, H. D., Least-Squares Absolute-Structure Refinement - Practical Experience and Ancillary Calculations. *Acta Crystallographica Section A* **1985**, *41*, 500-511.

Appendix II: Supplementary Information for Chapter 2

Stabilization of i-Motif Structures by 2'-β-Fluorination of DNA

AII.1 Differential Scanning Calorimetry Fitting Procedure

Each dataset was fit with linear folded and quadratic unfolded C_p baselines,[1] resulting in temperature dependent ΔC_p values over the experimental temperature range. The experimental data were buffer and baseline subtracted to give the excess heat capacities. For fitting excess heat capacities of intermolecular melting processes, it was necessary to compute the concentration of the monomeric state at all temperatures. An i-motif undergoing a two-state folding equilibrium involving tetrameric and monomeric states has total concentration of strands ($[M]_T$) equal to:

$$[M]_T = [M](T) + 4[T](T) \quad (1)$$

[M](T) is the concentration of free monomeric strands as a function of temperature, and [T](T) is the concentration of tetramer as a function of temperature. The coefficient of four accounts for the 4:1 stoichiometry of forming the tetramer from monomer strands. The temperature-dependent equilibrium constant for this process is:

$$K(T) = \frac{[T](T)}{\left([M](T)\right)^4} \quad (2)$$

which is related to the free energy of association according to:

$$K(T) = e^{\frac{-\Delta G(T)}{RT}} \quad (3)$$

where R is the ideal gas constant, and the change in free energy is:

$$\Delta G(T) = \Delta H(T) - T\Delta S(T) \quad (4)$$

where ΔG(T), ΔH(T), and ΔS(T) are the changes in Gibbs free energy, enthalpy, and entropy of folding at each experimental temperature. For thermodynamic processes with non-zero ΔC_p:

$$\Delta H(T) = \Delta H_o + \Delta C_p(T - T_o) \quad (5)$$

And

$$\Delta S(T) = \Delta S_o + \Delta C_p \ln \frac{T}{T_o} \tag{6}$$

Where ΔH_0 and ΔS_0 are the enthalpy and entropy of folding at the reference temperature, T_0. Substitution of eq. 2 into eq. 1 gives a fourth order polynomial:

$$[M](T) + 4K(T)\left([M](T)\right)^4 - [M]_T = 0 \tag{7}$$

Which can be solved numerically by substituting in the known value of $[M]_T$ and the test value of K(T) and taking the real, positive root as [M](T). The partition function is calculated as:

$$Z(T) = \frac{[M]_T}{[M](T)} = 1 + 4K(T)\left([M](T)\right)^3 \tag{8}$$

The fractional population of tetramers is then:

$$P_T(T) = \frac{K(T)\left([M](T)\right)^3}{1 + K(T)\left([M](T)\right)^3} \tag{9}$$

and the fractional population of monomers is:

$$P_M(T) = \frac{1}{1 + K(T)\left([M](T)\right)^3} \tag{10}$$

For i-motif melting equilibria comprising exchange between *n* folded states, the excess heat capacity at each experimental temperature is given by:

$$C_p(T) = \sum_{i=1}^{n} \Delta H_i(T) \times \frac{d}{dT} P_i + \sum_{i=1}^{n} P_i(T) \times \Delta C_{p,i}(T) \tag{11}$$

Where $\Delta C_{p,i}(T)$ is the difference in heat capacity between the *i*th folded state and the unfolded state at each temperature, and $\frac{d}{dT} P_i$ is the temperature derivative of the population of the *i*th state. This was calculated numerically according to:

$$\frac{d}{dT} P_i = a \frac{P_i(T(m+1)) - P_i(T(m))}{T(m+1) - T(m)} \tag{12}$$

Where a is the molecularity coefficient (0.25 for tetramer, 0.5 for dimer) and T(m) is the *m*th experimental temperature. The i-motif structures studied here required non-two state partition functions. H-1 to H-5 were analyzed using:

$$Z(T) = 1 + 4K_1(T)\left([M](T)\right)^3 + 4K_2(T)\left([M](T)\right)^3 \quad (13)$$

where the second and third terms correspond to two different tetrameric structures with folding equilibrium constants $K_1(T)$ and $K_2(T)$. H-6 was fit with a partition function assuming a dimer unfolding intermediate:

$$Z(T) = 1 + 2K_1(T)\left([M](T)\right)^1 + 4K_2(T)\left([M](T)\right)^3 \quad (14)$$

The data were fit by generating C_p profiles with initial baseline and thermodynamic parameters using eqs. 3-13 and minimizing residual sum of squares (RSS):

$$RSS(\xi)\sum_{j=1}^{N}(C_p^{\exp}(j) - C_{p,ij}^{fit}(\xi, j))^2 \quad (15)$$

Where N is the number of points in the experimental and fitted datasets and ξ= [ΔH_i, ΔS_i, a_u, b_u, c_u, m_f, b_f] where ΔH_i, ΔS_i are the folding enthalpy and entropy of the ith state, a_u, $b_{u,}$, c_u are the quadratic unfolded baseline parameters, and m_f, b_f are the linear folded baseline parameters.

AII.2 $T_{1/2}$ Data

Table AII.1: $T_{1/2}$ (°C) values for the (un)modified sequences at pH 5.0 and 0.2 °C/min.[a]

Code	Sequence (5'-3')	$T_{1/2}$	$\Delta T_{1/2}$
H-1	dTCCCCC	49.0	-
H-2	dTCC(fC)CC	9.2, 51.4	+2.4
H-3	dTC(fC)(fC)CC	23.6, 57.6	+8.6
H-4	dTCC(fC)(fC)C	21.7, 55.7	+6.7
H-5 [b]	dTC(fC)(fC)(fC)C	33.9	-
H-6 [b]	dTf(CCCCC)	33.8	-
HC-0	dTCCTTTTCCA	15.8	-
HC-1	dT(fC)CTTTTCCA	20.4	+4.6
HC-2	dT(fC)(fC)TTTTCCA	27.6	+11.8
HC-3	dT(fC)CTTTT(fC)CA	30.2	+14.4
HC-4	dT(fC)(fC)TTTT(fC)(fC)A	33.4	+17.6
HT-0	$d(CCCTAA)_3C_3$	57.2	-
HT-1	$d(C(fC)CTAA)_3C(fC)C$	65.0	+7.8
HT-2	$d(C(fC)(fC)TAA)_3C(fC)(fC)$	74.0	+16.8
HT-3	$d((fC)(fC)CTAA)_3(fC)(fC)C$	70.4	+13.2
HT-4	$d((fC)(fC)(fC)TAA))_3(fC)(fC)(fC)$	75.8	+18.6

[a]Oligonucleotide concentration was 4.6 µM in single strands for hexamers and 4.0 µM for HC and HT. $T_{1/2}$ data were calculated from UV-Visible spectroscopy thermal denaturation profiles where $T_{1/2}$ corresponded to the non-reversible melting temperature obtained at 0.2 °C/min. $\Delta T_{1/2}$ values were calculated for the main melting transitions relative to the respective unmodified strands. (-) indicates sequences for which no melting transition was detected, or an absence of $\Delta T_{1/2}$.

[b]Sequences H-5 and H-6 exhibit lower $T_{1/2}$ values compared to the control (H-1). This was due to the melting of a dimeric structure rather than an i-motif structure. As a result, $\Delta T_{1/2}$ cannot be calculated since the comparison would H-1 will be invalid.

AII.3 CD Spectra

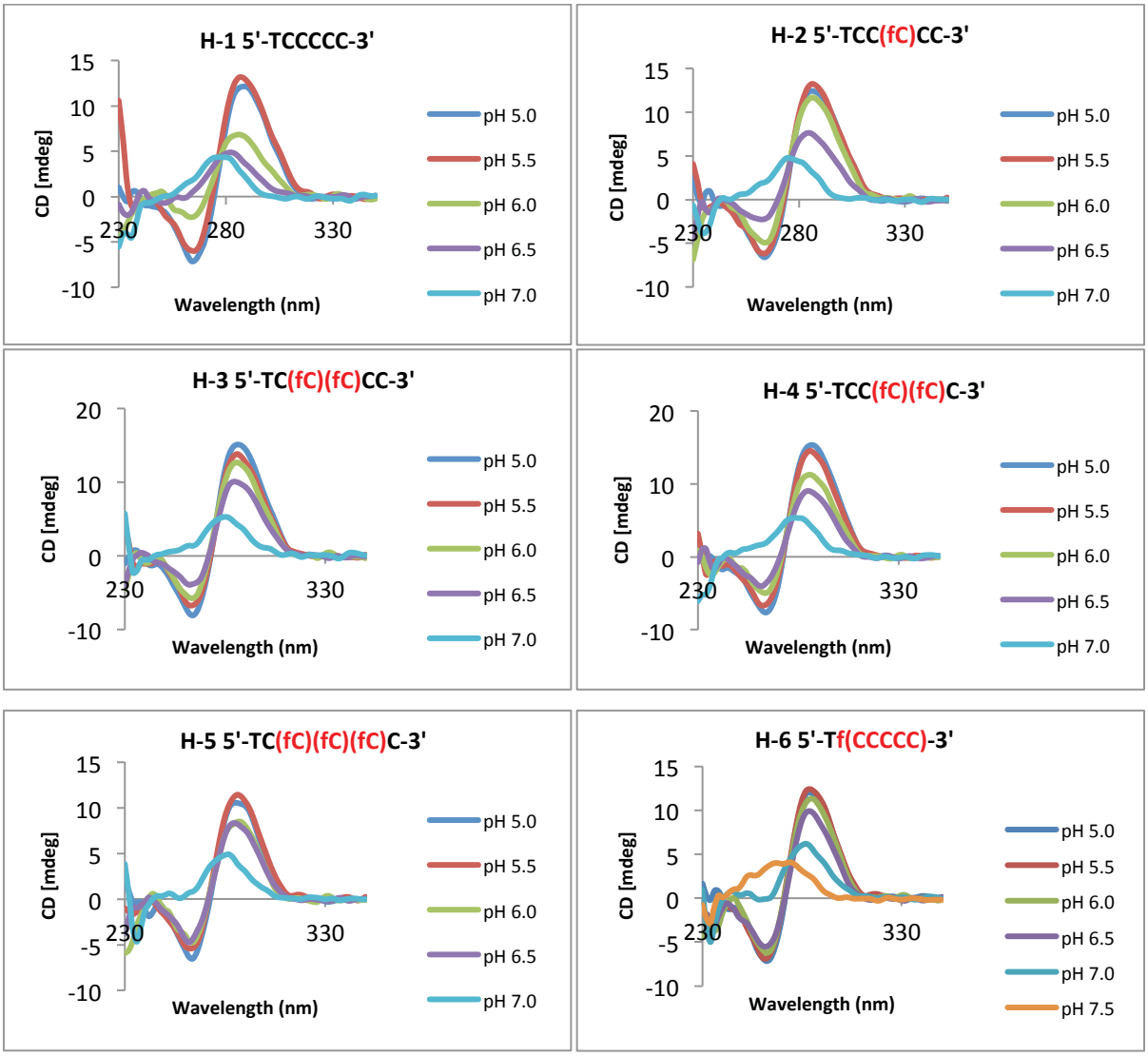


Figure AII.1: CD spectra of hexamer sequences at pH 5.0-7.5. The strand concentration was 30 µM and the buffer was 10 mM sodium phosphate. Spectra were recorded at 5 °C.

AII.4 ^{1}H-NMR Spectra for Hexamer Sequences

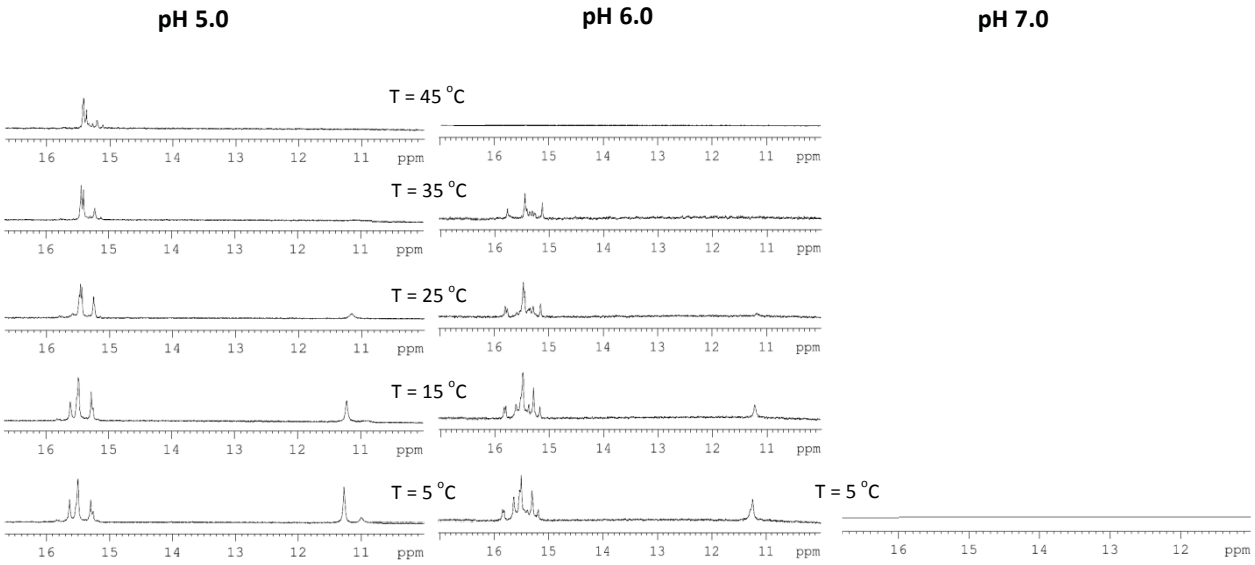


Figure AII.2-a: ^{1}H-NMR melting experiments for H-2 5'-TCC(fC)CC-3' (2.0 mM). At pH 7.0 no imino peaks were observed at 5 °C.

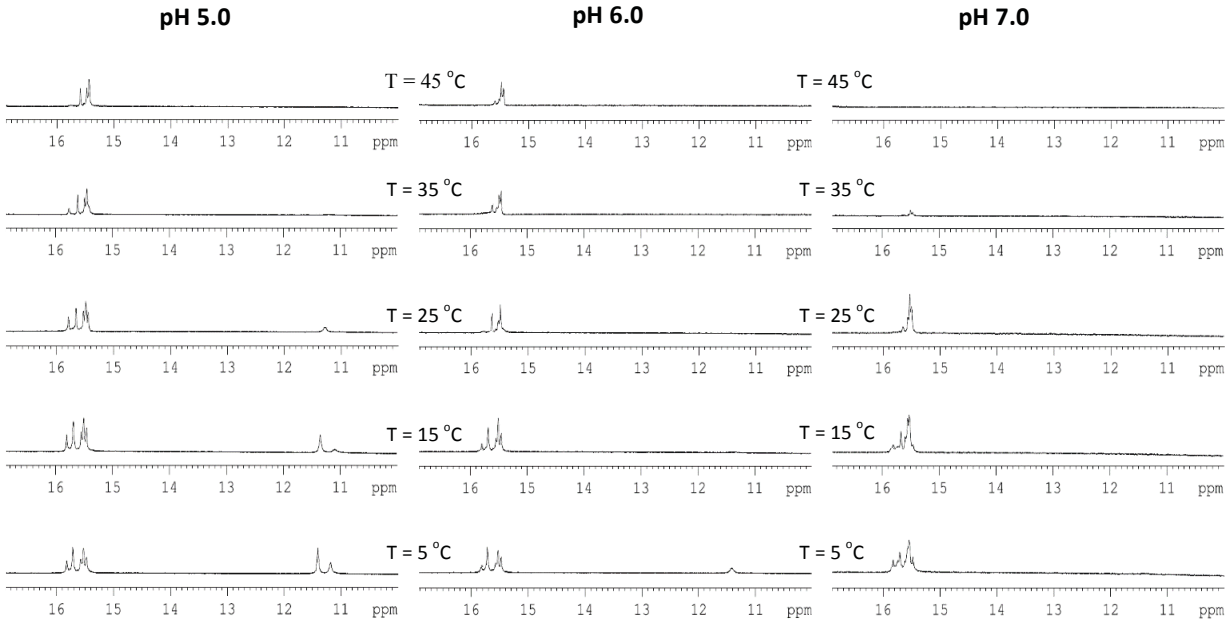


Figure AII.2-b: ^{1}H-NMR melting experiments for H-3 5'-TC(fC)(fC)CC-3' (2.6 mM).

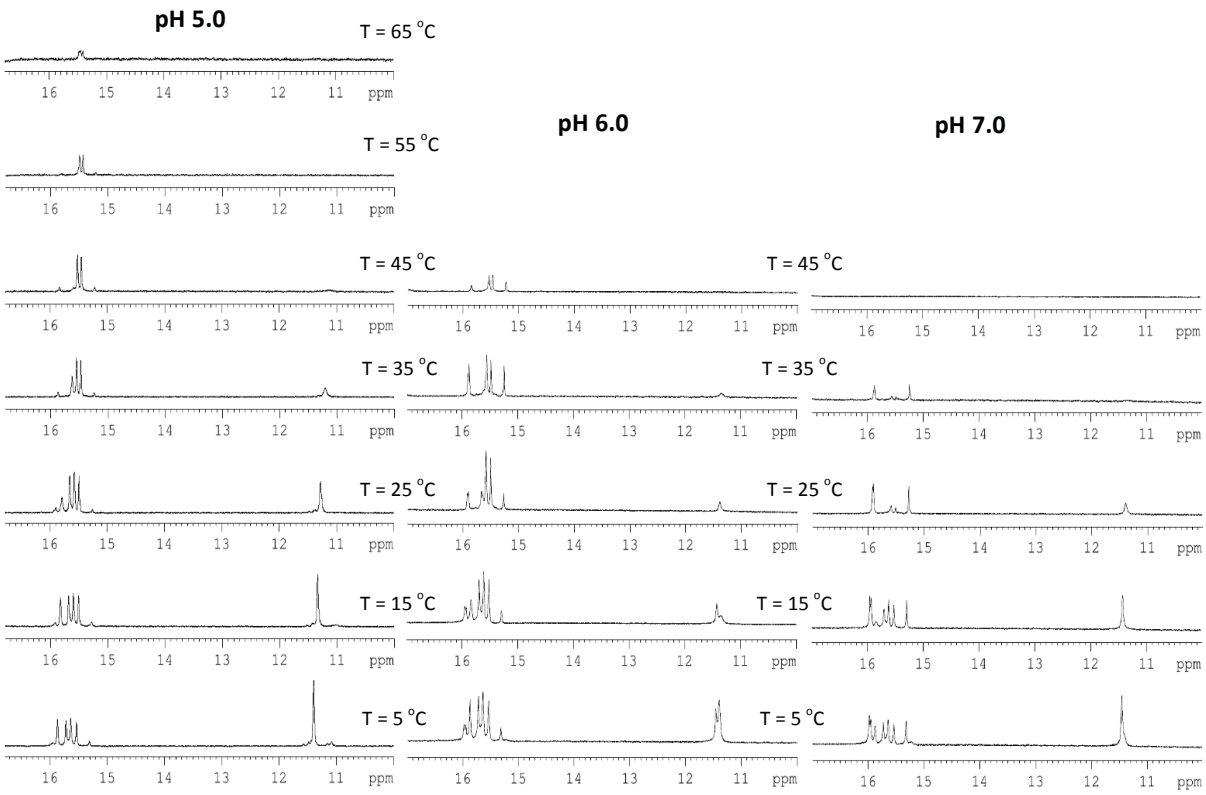


Figure AII.2-c: ^{1}H-NMR melting experiments for H-4 5'-TCC(fC)(fC)C-3' (2.53 mM).

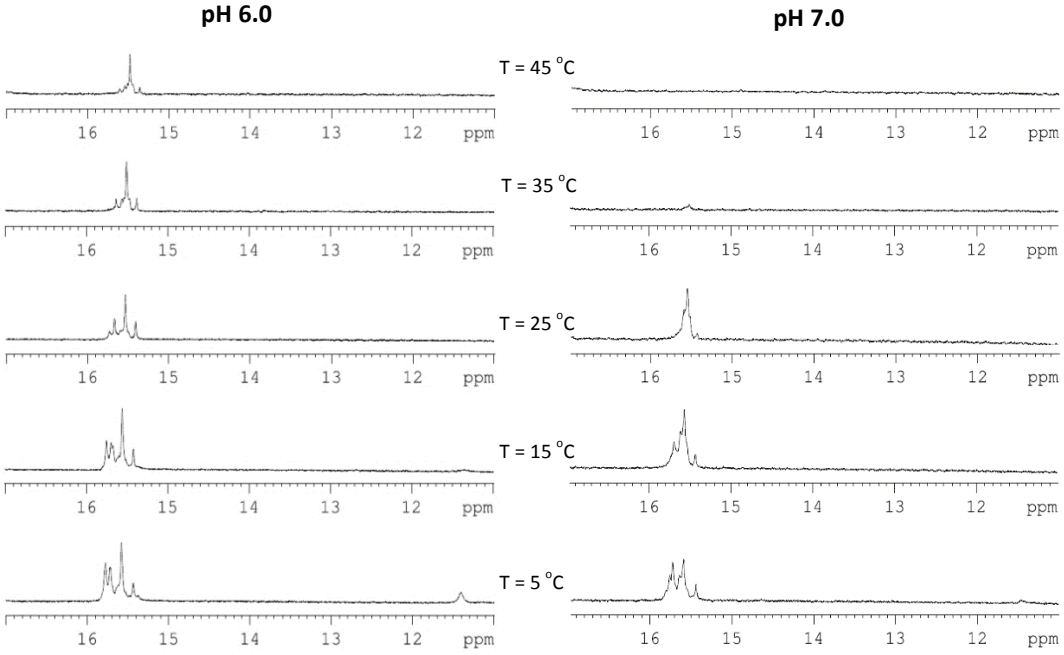


Figure AII.2-d: ^{1}H-NMR melting experiments for H-5 5'-TC(fC)(fC)(fC)C-3' (2.08 mM).

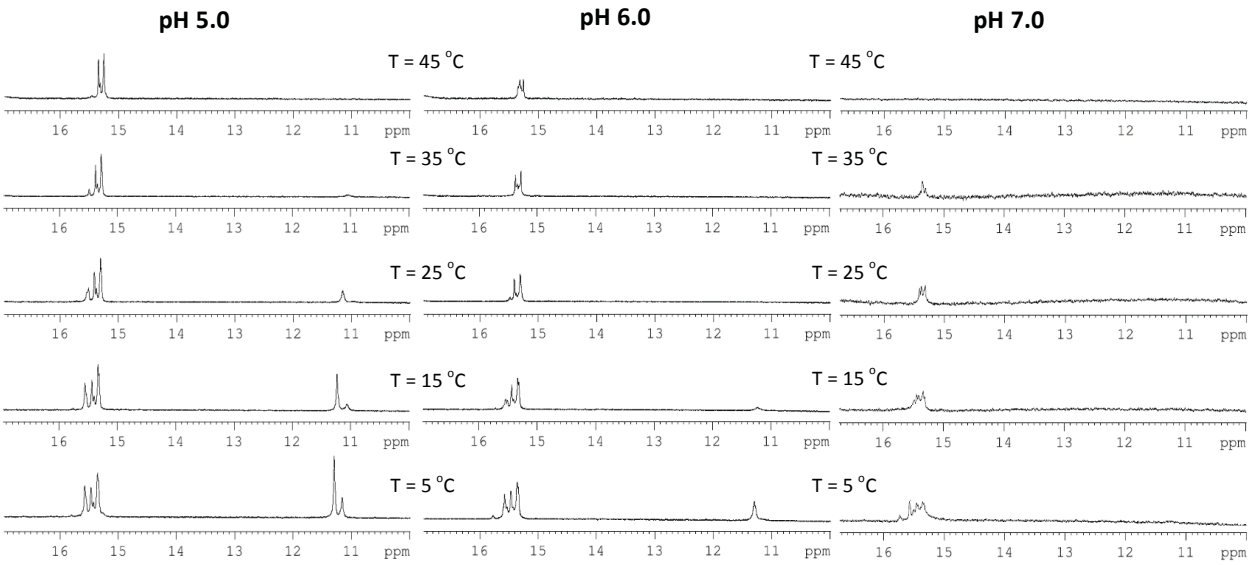


Figure AII.2-e: ^{1}H-NMR melting experiments for H-6 5'-Tf(CCCCC)-3' (2.1 mM).

AII.5 ^{19}F-NMR Spectra for Hexamer Sequences

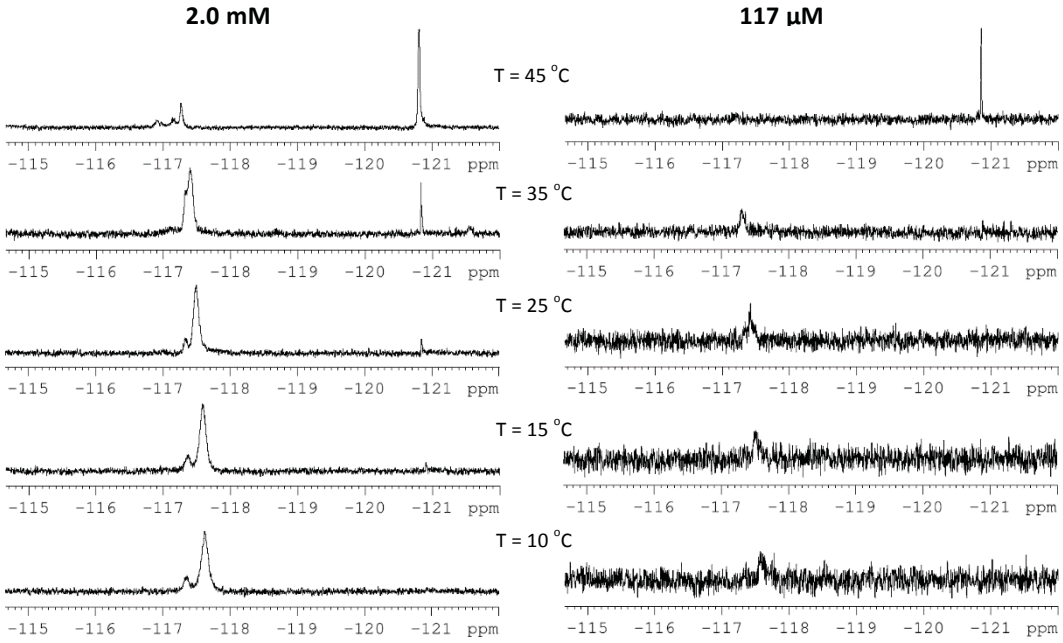


Figure AII.3-a: ^{19}F-NMR melting experiments for H-2 5'-TCC(fC)CC-3', pH 5.0.

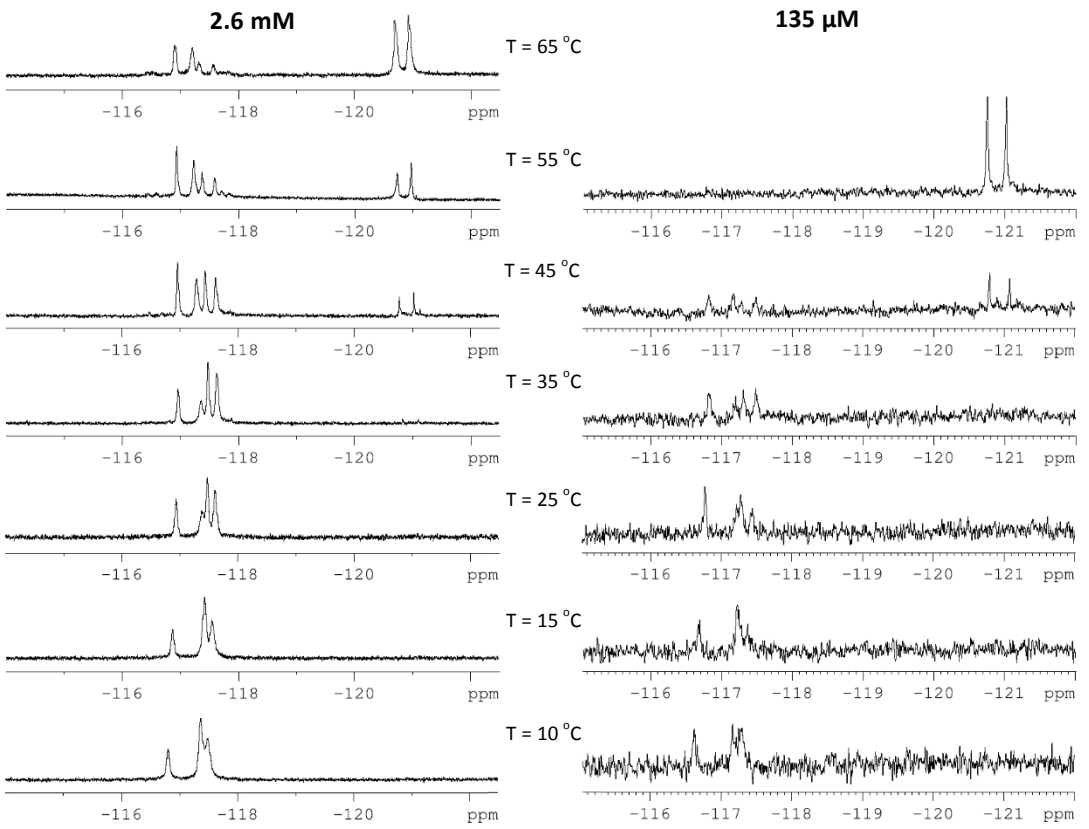


Figure AII.3-b: ^{19}F-NMR melting experiments for H-3 5'-TC(fC)(fC)CC-3', pH 5.0.

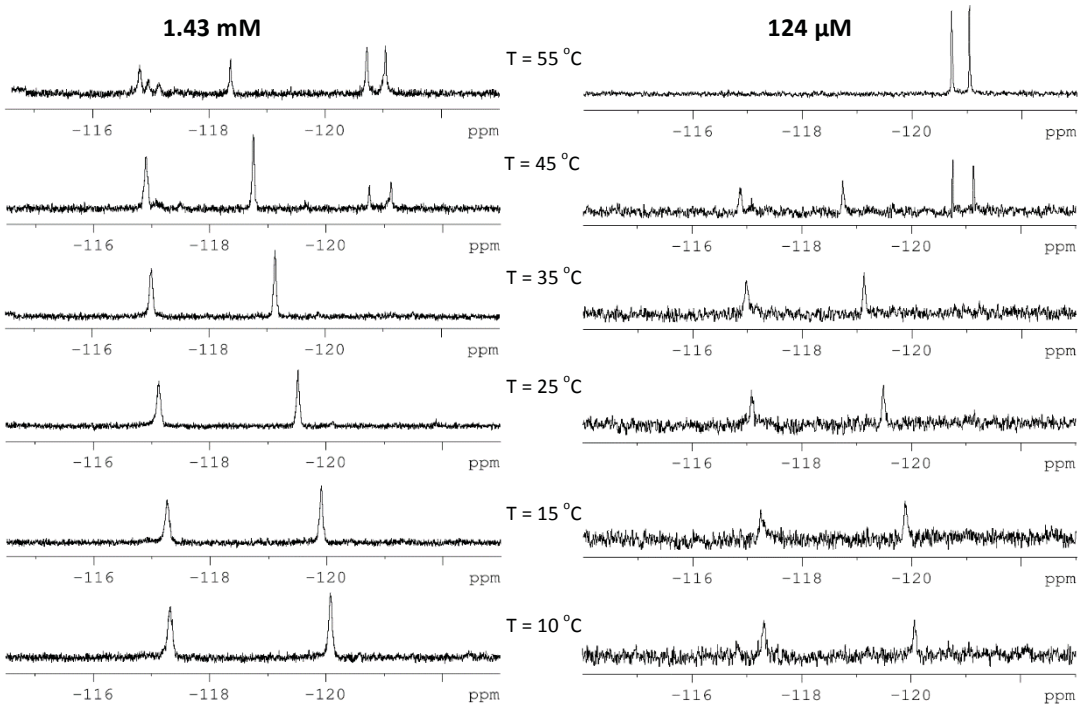


Figure AII.3-c: ^{19}F-NMR melting experiments for H-4 5'-TCC(fC)(fC)C-3', pH 5.0.

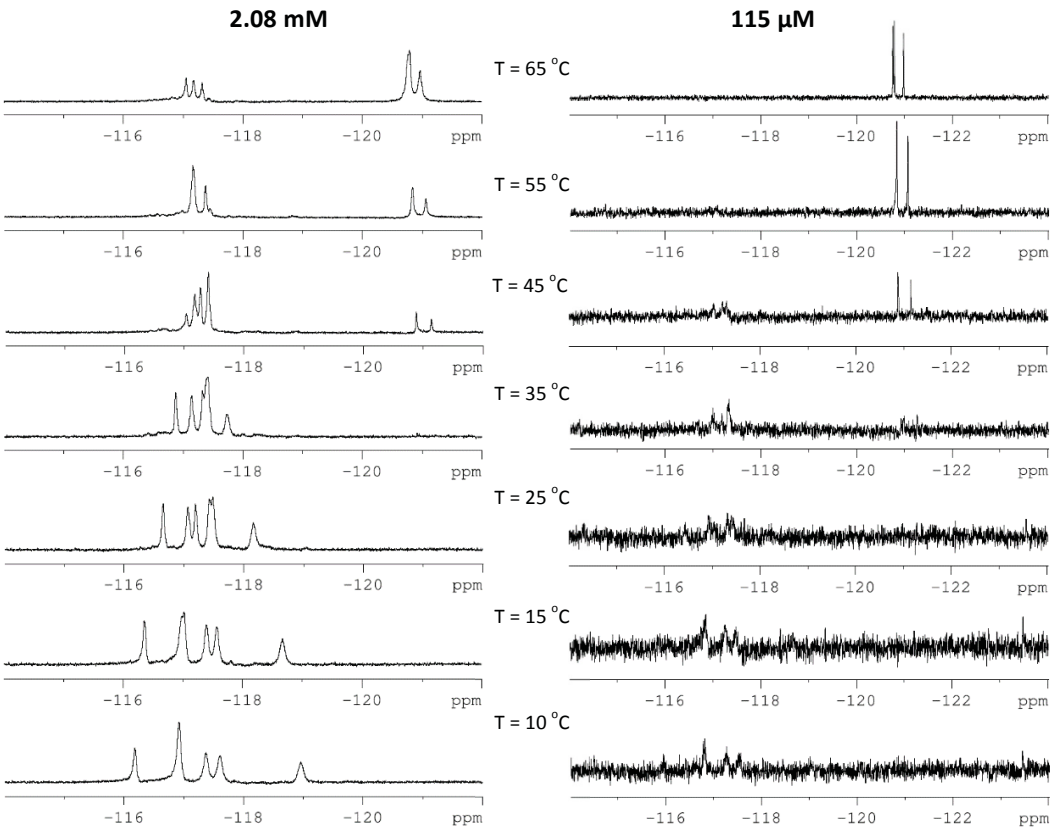


Figure AII.3-d: ^{19}F-NMR melting experiments for H-5 5'-TC(fC)(fC)(fC)C-3', pH 5.0.

AII.6 DSC Data

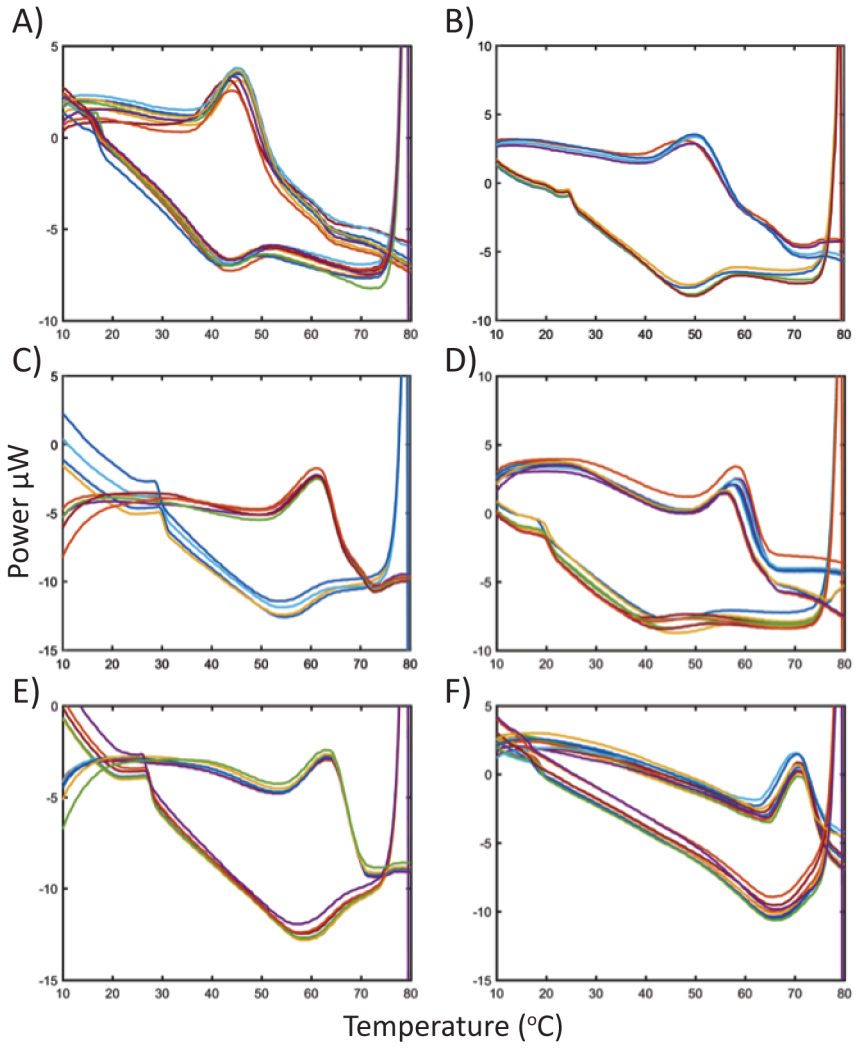


Figure AII.4: Raw DSC power data for unmodified and modified tetrameric i-motif structures. A) unmodified H-1, B) H-2, C) H-3, D) H-4, E) H-5, and F) H-6. The forward and reverse scans showed minimal hysteresis at 150 µM concentrations, in contrast to the UV-Vis melts at 4.6 µM.

AII.7 NMR Data

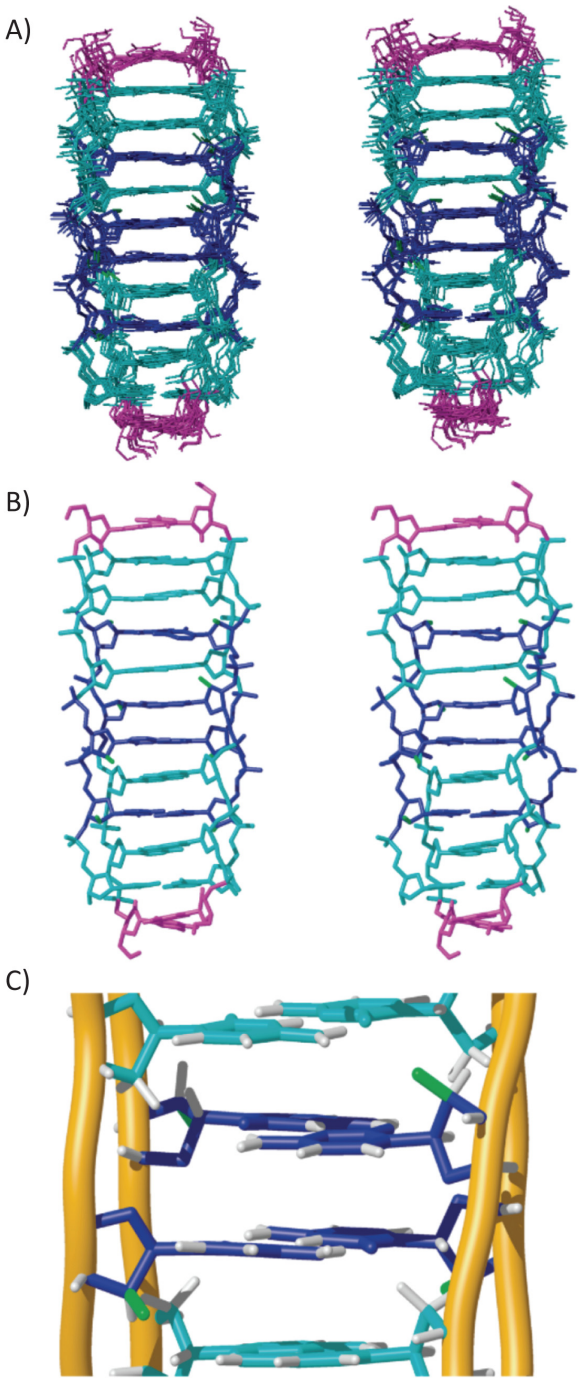


Figure AII.5: Tetrameric structure of H-4 5'-TCC(fC)(fC)C-3' at acidic pH. A) Stereoview of the ensemble of the 10 calculated structures. B) Stereoview of the average structure. C) Detail of the stacking interaction between $C{\cdot}C^{+}$ base pairs. Note that fluorine atoms are exposed to the solvent, pointing towards the major groove. Color code: dC and 2'F-araC are shown in cyan and blue, respectively; thymines are shown in magenta; fluorine atoms are shown in green. PDB code: 2N89

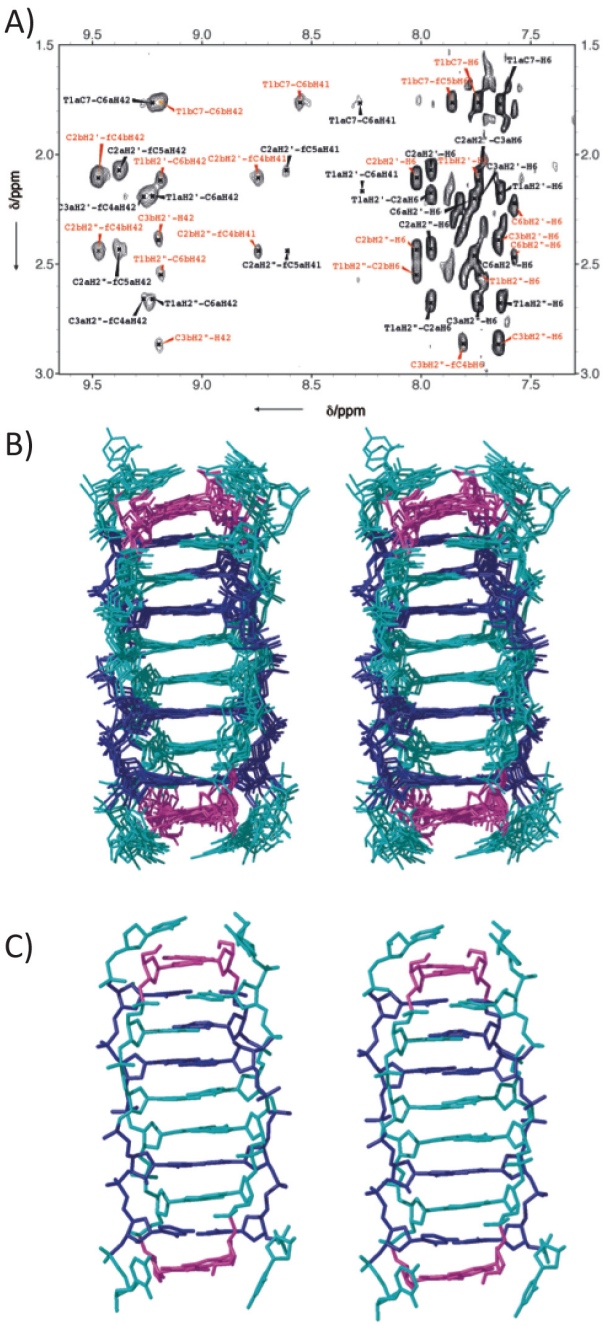


Figure AII.6: NMR assignment of the major species of H-4 at neural pH. A) Amino-H2'/2" region of the NOESY spectra (mixing time = 100 ms, T = 5 ^{o}C). Cross-peaks labelled in black correspond to the major species at acidic pH; cross-peaks labelled in red correspond to the major species at neutral pH. These NOEs together with those observed in the imino-amino region (Figure 2.20) clearly show that the major species at neutral pH is an i-motif. B) and C) Model structures of the major species of H-4 at neutral pH. Although a complete assignment of the neutral species could not be carried out, this model was consistent with the NOE connectivities observed. B) Stereoview of 10 structures calculated by molecular dynamics methods. C) Stereoview of a representative structure. Color code: dC and 2'F-araC are shown in cyan and blue, respectively; thymines are shown in magenta; fluorine atoms are shown in green.

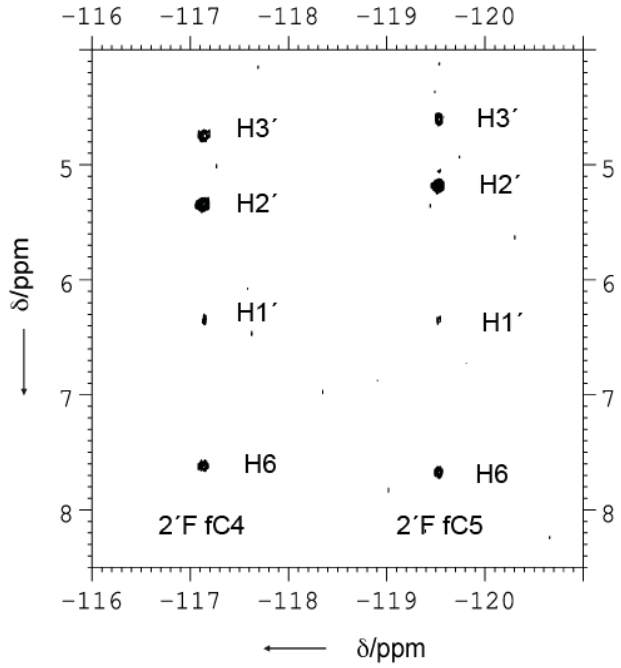


Figure AII.7: ^{19}F-^{1}H HOESY spectrum of (H-4) 5'-TCC(fC)(fC)C-3', pH 5.0, T = 25 ^{o}C (mixing time 200 ms).

Table AII.2: Assignment list of H-4 at pH 5.0/6.0, T = 5 ^{o}C, [H-4] = 2.5 mM.

	H3	H42	H41	H6	H5/Me	H1'	H2'	H2''	H3'	H4'	H5'	H5''
T1	11.46	-	-	7.63	1.75	6.34	2.17	2.66	4.83	n.a.	n.a.	
C2	15.88	9.34	8.32	7.95	5.90	6.58	2.06	2.42	4.82	n.a.	n.a.	
C3	15.65	9.25	8.47	7.73	5.88	6.40	2.18	2.66	4.89	n.a.	n.a.	
fC4	15.54	9.26	8.52	7.61	5.95	6.33	-117.13	5.37	4.75	n.a.	n.a.	
fC5	15.73	9.37	8.62	7.70	5.83	6.35	-119.52	5.16	4.57	n.a.	n.a.	
C6	n.o.	9.22	8.26	7.74	5.91	6.16	2.21	2.47	4.30	n.a.	n.a.	

n.o: not observed

n.a: not assigned

Note: Fluorine signals were assigned from the ^{19}F-^{1}H HOESY spectra of H-4, pH 5.0, T = 25 ^{o}C (mixing time 200 ms) (Figure AII.7).

Table AII.3: Experimental constraints and calculation statistics of H-4.

Experimental distance constraints		
Total number	284	
intra-residue	100	
sequential	16	
range > 1	168	
Intra-subunit	116	
Inter-subunit	168	
RMSD (Å)		
bases	0.5±0.1 Å	
backbone	0.8±0.3 Å	
all heavy atoms	0.9±0.2 Å	
Residual violations	Average	Range
Sum of violation (Å)	2.5	1.8 - 3.3
Max. violation (Å)	0.3	0.2 - 0.5
NOE energy[#] (kcal/mol)	5.1	4.5 - 7.1
Total energy (kcal/mol)	-2583	-2441 - -2694

[#] K_{NOE} = 20 kcal/(mol .Å^2)

Table AII.4: Average dihedral angles and order parameters of H-4.

	Pseudo-rotation		**α**		**β**		**Y**		**δ**		**ε**		**ζ**		**X**	
	Phase	Amplitude	Average	Order parameter	Average	Order parameter	Average	Order parameter	Average	Order parameter	Average	Order parameter	Average	Order parameter	Average	Order parameter
T1	33	34	-	-	-	-	67	0.8	84	1	171	0.6	-	-	-114	1
C2	27	36	-55	0.6	180	1	91	0.7	85	1	-167	1	-84	0.5	-120	1
C3	23	33	-105	0.5	179	1	109	0.7	90	1	-135	0.8	-87	0.9	-123	1
fC4	134	33	-101	0.8	113	0.6	164	0.7	124	1	-167	0.6	-72	0.9	-122	1
fC5	118	39	-133	0.5	150	0.8	-169	1	110	1	-174	0.6	-64	0.5	-118	1
C6	58	43	-119	0.8	170	1	-169	1	80	1	-	-	-71	0.5	-115	1

AII.8 References

1. Privalov, G.; Kavina, V.; Freire, E.; Privalov, P.L., Precise Scanning Calorimeter for Studying Thermal-Properties of Biological Macromolecules in Dilute-Solution. *Anal. Biochem.* **1995**, ***232***, 79-85.