Effects of High Pressure on DNA and its Components

A thesis presented in partial fulfilment of the requirements for the degree of

PhD
in
Bio Physics

at Massey University, Manawatū,
New Zealand.

Christopher Paul Lepper

Abstract

There have been many speculations for the environment in which life originated but it has still yet to be determined what environmental chemical and physical conditions were necessary for the evolution of self-replicating chemical systems. While it has been determined that DNA, RNA and their components are chemically unstable at high temperatures, there has currently been only a small number of studies into the role of high pressures on the chemical and physical stabilities.

High-pressure NMR spectroscopy has been used here to study the effects of high temperatures/pressures on the chemical stability of DNA and its components. This has been done with the use of a specialised commercial high-pressure NMR cell capable of withstanding pressures up to 250 MPa. In addition to this, a custom safe handling apparatus and pump system was developed for the operation of this cell.

Studies into the effects of high pressures on the rate of hydrolysis of cytosine and cytidine at 100 °C were performed by measuring the rates of hydrolysis with time under various pressure conditions. These results have shown that the rates of hydrolysis of cytosine and cytidine increase considerably with pressure.

The effects of high pressure on the physical stability of DNA were determined by performing dissociation (melting) experiments on several different DNA sequences under multiple pressure conditions. It was found that the melting point of a small DNA hexamer decreased slightly with pressure whereas the melting points of larger dodecamers increased overall with pressure. It was also found that the melting point of an i-motif structure decreased with increasing pressure.

The effects of high pressure on the chemical stability of cytosine were again studied, this time for cytosine residues within both single- and double-stranded DNA. DNA samples for bacteriophage ΦX174 were incubated under various temperature and pressure conditions. Results for these studies have yet to be determined as the incubated DNA is yet to be sequenced.

It has been discovered that high pressures have a negative effect on the chemical stability of DNA constituents while having an overall small positive effect on the physical stability of DNA.

Acknowledgements

Firstly I would like to thank my supervisors, Prof. Geoffrey Jameson, Dr. Patrick Edwards and Prof. Martin (Bill) Williams, for their fantastic help and support throughout this project. I would like to offer a special thanks to Dr. Vyacheslav Filichev, Dr. Mark Patchett and Assoc. Prof. Murray Cox; this project would have not been possible without their expertise and help. I would also like to thank the Massey University IFS Engineering work shop who fabricated a large portion of the equipment for the pump system used in this study.

This study has been made possible with a grant from the Marsden fund provided by the Royal Society of New Zealand. This grant has provided funding for both the project and my living expenses.

This work has been presented at the Sixth International Conference on Advanced Materials and Nanotechnology (AMN6) in Auckland (11th-15th February 2013) and at the Second Physics of Biology Meeting in Geneva, Switzerland (26th-28th November 2013. Attendance at these conferences would not have been possible without funding from Massey University Institute of Fundamental Sciences and the New Zealand Institute of Chemistry, Manawatū Branch.

Lastly I would like to thank my partner Tina and my family who have continually supported me in my never ending quest to remain at university.

Contents

Abstract i

Acknowledgements ii

Figures and Tables viii

Figures viii

Tables xvi

List of Nomenclature and Abbreviations xix

Nomenclature xix

Abbreviations xxvi

Chapter 1 - Introduction 1

1.1 Overview 1

1.2 Popular Theories 2

1.2.1 Cold-Start Theories 3

1.2.2 Warm-Start Theories 3

1.2.3 Hot-Start Theories 5

1.3 Effects of Extreme Conditions on Biological Molecules 6

1.3.1 High-Temperature Effects 6

1.3.2 High-Pressure Effects 7

1.4 Aims of this Thesis 8

1.4.1 Development of a High-Pressure System 9

1.4.2 Chemical Stability of Individual DNA Bases 9

1.4.3 Physical Stability of DNA Structures 10

1.4.4 Chemical Stability of DNA 11

1.5 Summary of Thesis Structure 11

Chapter 2 - The High-Pressure System 13

2.1 High-Pressure NMR Cell 13

2.1.1 Fused Silica Capillaries 13

2.1.2 Metallic Pressure Vessels 14

2.1.3 Diamond Anvil Cells 15

2.1.4 Large-Volume Tubes 16

2.1.5 The Daedalus Innovations Cell 17

2.2 Safe-Handling Apparatus for the High-Pressure NMR Cell 19

2.3 High-Pressure Pump System 22

2.3.1 Initial Development 22

2.3.2 Upgrades to High-Pressure System 24
2.3.3 Safety 26
2.3.4 Choice of Hydraulic Fluid 26
2.4 High-Pressure Reaction Vessel 26
2.5 Buffer Correction 28
2.6 Problems and Failures 30
2.6.1 Tuning Issues 30
2.6.2 Cell Failure 31
Chapter 3 - Cytosine Deamination 33
3.1 Introduction 33
3.1.1 Earlier Studies 33
3.1.2 Consequences 35
3.1.3 High-Pressure Effects 36
3.2 Materials and Methods 36
3.2.1 Cytosine/Cytidine Solutions 36
3.2.2 Incubation 40
3.2.3 Hydrolysis Measurements 41
3.3 Results 43
3.3.1 Cytosine Hydrolysis 45
3.3.2 Cytidine Hydrolysis 48
3.4 Discussion 49
3.5 Conclusions 53
3.6 Future Work 53
3.6.1 Salts 54
3.6.2 pH Dependence 54
3.6.3 Molecular Crowding and Effects of Solutes 55
Chapter 4 - Physical Stability of Hexamers 57
4.1 Introduction 57
4.1.1 Effects of High Pressure on Protein Stability 57
4.1.2 Effects of High Pressure on DNA/RNA Stability 58
4.2 Materials and Methods 59
4.2.1 Sample Preparation 59
4.2.2 DNA Analysis 60
4.2.3 Spectral Assignment 61

4.2.4 Spectral Assignment of 5'-CAT ATG-3' 68
4.2.5 Melting Experiments 74
4.2.6 Circular Dichroism Experiments 74
4.3 Results 74
4.3.1 Data Gathering and Processing 74
4.3.2 Data Fitting 77
4.3.3 Circular Dichroism Results 80
4.4 Discussion 82
4.5 Conclusions 84
4.6 Future Work 84
4.6.1 Longer Sequences 85
4.6.2 G-C content 85
4.6.3 RNA 85
4.6.4 Salts 86
Chapter 5 - Physical Stability of Dodecamers 87
5.1 Introduction 87
5.2 Materials and Methods 88
5.2.1 Sample Preparation 88
5.2.2 DNA Analysis 90
5.2.3 Spectral Assignment of Sequence 123 91
5.2.4 Melting Experiments 98
5.2.5 Circular Dichroism Experiments 98
5.3 Results 98
5.3.1 Data Gathering and Processing 98
5.3.2 Data Fitting 105
5.3.3 Circular Dichroism Results 111
5.4 Discussion 113
5.4.1 Sequence 121 (16.7 % G-C Content) 113
5.4.2 Sequence 122 (33.3 % G-C Content) 116
5.4.3 Partial Molar Isothermal Compressions at Infinite Dilution 117
5.4.4 Sequence 123 (50.0 % G-C Content) 120
5.4.5 Sequence 124 (66.7 % G-C Content) 121
5.4.6 Sequence 125 (83.3 % G-C Content) 123
5.4.7 Influences of Pyrimidines and Purines 125

5.5 Conclusions 127
5.6 Future Work 128
5.6.1 Different Sequences 128
5.6.2 RNA 129
Chapter 6 - Physical Stability of i-Motifs 131
6.1 Introduction 131
6.1.1 Alternate DNA Structures 131
6.1.2 Tetramers and Their Importance 134
6.1.3 Effects of High Pressures on G-Quadruplexes and i-Motifs 134
6.2 Materials and Methods 136
6.2.1 Sample Preparation 136
6.2.2 NMR Spectroscopy 137
6.2.3 Circular Dichroism Spectroscopy 138
6.3 Results 139
6.3.1 Spectral Assignment 139
6.3.2 Data Gathering 139
6.3.3 Effects of pH and Pressure on i-Motif Stability 141
6.3.4 Circular Dichroism results 143
6.4 Discussion 144
6.4.1 i-Motif in Phosphate Buffer 144
6.4.2 i-Motif in Citrate Buffer 146
6.5 Conclusions 149
6.6 Future Work 150
6.6.1 Extra pH Measurements 150
6.6.2 Salts 150
6.6.3 Molecular Crowding 151
Chapter 7 - Chemical Stability of Cytosine in ϕX174 153
7.1 Introduction 153
7.2 Materials and Methods 154
7.2.1 Sample Preparation 156
7.2.2 Incubation 158
7.2.3 PCR Amplification 158
7.2.4 Sample DNA Integrity Checks 160
7.3 Next-Generation Sequencing 160

7.3.1 Sequencing by Synthesis 161
7.3.2 The Illumina Process 162
7.4 Results 164
7.4.1 Pre Amplification Gel Electrophoresis 164
7.4.2 Post Amplification Gel Electrophoresis 166
7.4.3 DNA Sequencing 168
7.5 Discussion 168
7.5.1 283.2 K Incubation 168
7.5.2 313.2 K Incubation 169
7.5.3 333.2 K Incubation 169
7.6 Conclusions 170
7.7 Future Work 170
7.7.1 Data Processing 171
7.7.2 Other Conditions 171
Chapter 8 - Conclusions 173
Bibliography 175
Appendices 185
Appendix one – Cytosine/Cytidine Hydrolysis Data 185
Appendix Two – Hexamer Melting Data 186
Appendix Three – Dodecamer Melting Data 187
Spectral Assignment 187
CD Spectroscopy Results 192
Melting Results 197
Appendix Four – i-Motif Melting Data 200
Appendix Five –Final Concentrations of Amplified DNA 202
Appendix Six – High-Pressure System Operational Manual 204
Appendix Seven – Digital Appendices 228

Figures and Tables

Figures

Figure 1.1: The Fly Geyser, Fly Ranch Nevada. This warm alkaline spring provides suitable conditions for the growth of red and green algae on the calcium mineral deposits [20]. Image obtained from http://exotic-place.com/wp-content/uploads/2013/12/Fly-Geyser-Nevada3.jpg. 4

Figure 1.2: Tufa Formations at Mono Lake, California. These formations are created by calcium deposits from submarine alkaline springs. A receding water level has now exposed this formation [21]. Image obtained from http://img.phombo.com/img1/photocombo/2711/cache/Tufa_Towers_with_ Sierra_Mountains_at_Twilight_Mono_Lake_California_display.jpg. 4

Figure 1.3: A Submarine Hydrothermal Vent. This black smoker provides energy and minerals for the surrounding life forms. Image obtained from http://www.mpi-bremen.de/Binaries/Binary5561 /black_Smoker1.jpg. 5

Figure 1.4: Equation for the Hydrolysis of Cytosine to Uracil. 10

Figure 2.1: Fused Silica Capillary Cell. The capillary is bent back and forth several times to increase the sample volume within the RF field. it is then placed within a standard NMR tube for protection. Image adapted from Conradi [41]. 14

Figure 2.2: Metallic Pressure Vessel. This cell is constructed from a metal pressure vessel (1) containing the sample tube (2) which is placed inside the RF coil (3). The vessel is sealed from the bottom with a plug (4) and pressure is provided from an external pressure source connected to the top of the cell (5). Image adapted from Conradi [41]. 15

Figure 2.3: Diamond Anvil Cell. The sample is placed in the sample region (1) between the two diamonds (2). A metallic gasket (3) holds the sample in place. The RF coil is situated around this middle region of the cell. Pressure is applied by tightening the screws (4) which connect the backing plates (5). Image adapted from Conradi [41]. 16

Figure 2.4: High-Pressure NMR Tube. The sample is placed in the tube (1) which is connected to a manifold (2). The cell is pressurised by a pump connected to the high-pressure connection point (3). Image adapted from Daedalus Innovations [45]. 17

Figure 2.5: The Daedalus Innovations Cell. (a) A schematic of the Daedalus cell. The ceramic tube is secured in the cell manifold by the manifold base. A nitrile O-ring (TS01) provides the seal while the metallic tube seat (TCSN-M) ensures correct seating of the tube. Pressure is provided via a high-pressure connection at the top of the cell. Image adapted from Daedalus Innovations. (b) The cell manifold and the completed cell. Image obtained from http://www.daedalusinnovations.com /apparatus/high-pressure.html. 19

Figure 2.6: Safe-Handling Apparatus for the High-Pressure NMR Cell. This schematic outlines the important features of the safe-handling apparatus. Further information on these features is given in the text. (1) The high-pressure cell. (2) Aluminium shaft. (3) Shaft sleeve. (4) Alignment bracket. (5) Main arm. (6) Secondary arm. (7) High-pressure pump. (8) Wall-mounting bracket. (9) Locking arm. (10) High-pressure tether. 22

Figure 2.7: Schematic of High-Pressure Pump. The high-pressure NMR cell is connected to the cell outlet valve via the high-pressure tether connected at the cell outlet. 24

Figure 2.8: High-Pressure Connection. A collar is screwed onto the end of the high-pressure tubing. A gland nut is then threaded over the collar and screwed into the component. This forces the coned end of the tubing in the orifice of the component creating a seal. This connection may be disassembled and reassembled indefinitely. The threads of the gland are right-hand while the threads of the collar and tubing are left-hand to prevent rotation of the collar during assembly. Image adapted from HiP [48]. 24

Figure 2.9: Schematic of High-Pressure Pump Upgrade. The external outlet valve connects to the high pressure reaction vessel via a high-pressure tether..... 25

Figure 2.10: High-Pressure Reaction Vessel. (a) A cross section of the reaction vessel showing the cap (1), body (2), vent hole (3), O-ring (4) and backup ring (5). (b) The reaction vessel in its frame. This shows the high-pressure tether connection (6), the isolation valve (7) tubing to connect the valve to the underside of the reaction vessel (8) and the high-pressure plug (9).. 28

Figure 2.11: Cross Section of Cell Manifold Base as Seen From Below. The dashed lines indicate the flattened surfaces which allow for a spanner to grip this section for cell assembly..... 31

Figure 2.12: Broken High-Pressure Tube. Base of tube, 0 mm. The highest point of the breakage, 46 mm. Oil level when immersed, 70 mm. Top of tube, 92 mm. 32

Figure 3.1: DNA/RNA Bases. Adenine (A), cytosine (C), guanine (G), thymine (T) and uracil (U). The alternate bases xanthine, hypoxanthine, diaminopyrimidine, isocytosine, isoguanine and diaminopurine are also shown. 34

Figure 3.2: High-Pressure NMR Cell in Oil Bath. This figure shows the depth to which the high-pressure NMR cell was placed in the oil bath. It can be seen that the entire sample (red) within the cell is below the oil level. 41

Figure 3.3: Cytosine/Cytidine Hydrolysis. a) Cytosine to uracil and b) cytidine to uridine indicating hydrogens 5H and 6H. c) An example spectrum of cytosine during hydrolysis. The peaks from hydrogens 5H and 6H from both cytosine and uracil have been labelled: for cytosine, C5 and C6, and for uracil, U5 and U6. These peaks were used to determine the rate of hydrolysis. 42

Figure 3.4: Mechanism of Cytosine Deamination. This figure outlines the mechanism by which cytosine or cytidine undergoes deamination. A resonance shift leaves a positive charge on C4 making it susceptible to attack from a water molecule. The amino group then leaves as an

NH_3.The N3 then takes the proton from the neighbouring OH. A final resonance shift restores the molecule to its final state as uracil. Alternative mechanisms have been proposed including catalytic activity from the buffer [58]. However, these mechanisms have not been confirmed and still share the same rate-limiting step. As such, the mechanism outlined in this figure will suffice for understanding the rate relationship. 44

Figure 3.5: ln N_f versus Time for Cytosine Hydrolysis at pH 7. The results of each pressure condition are shown. 46

Figure 3.6: k_{obs} versus pH. k_{obs} values at pH 6.0, 7.0 and 8.0 for both 0.1 MPa and 150 MPa. Connecting lines are as a visual reference only. 47

Figure 3.7: ln k_{obs} versus Pressure for Cytosine. From the gradient of this plot, the reaction volume has been calculated to be -11.7 ± 1.2 cm^3 mol^{-1}. 47

Figure 3.8: ln N_f versus Time for Cytidine Hydrolysis at pH 7. 48

Figure 3.9: Overlapping Peaks. An example of overlapping proton peaks where the midpoint of the proton signal was estimated. In this instance the uracil signal could possibly be offset by the overlap. 51

Figure 4.1: Nucleobase Labelling. Proton assignment diagram for all bases including labelling for the deoxyribose sugar. n' refers to a proton on the 2-deoxyribose group while n" refers to the second of two hydrogens attached to the same carbon. 63

Figure 4.2: Observed Proton Spectral Ranges. The spectral ranges over which specific proton signals occur as observed over the course of this study. Conveniently, C-6H signals can be identified as doubles while A-2H exhibit no NOE's with 2'H/2"H protons in a NOESY spectrum. This allows for the clear identification of the A-8H, G-8H, T-6H and T-5CH_3 signals. These results compare well with results seen by Nielsen *et al.* [77]. 64

Figure 4.3: Intraresidue (i) and Sequential (s) Connectivities Between Adjacent Nucleotides. The ^{1}H – connectivities d_i(6H/8H→1'H) and d_s(1'H→6H/8H) (red arrows), d_i(6H/8H→2H')/d_i(6H/8H→2H") and d_s(2'H→6H/8H)/d_s(2"H→6H/8H) (blue arrows) and d_s(6H/8H→5CH_3) (green arrows) are used for sequential assignments of non-labile protons in B-DNA. Image adapted from Wüthrich [78]. 65

Figure 4.4: Example of Cross Peaks and Their Links. Examples of the NOE cross peaks and their links in the anomeric (6.5 to 5.5 ppm)/base proton (8.5 to 7.0 ppm) region. Horizontal lines indicate i – s links and vertical lines indicate s – i links. 66

Figure 4.5: Examples of Partial Sequence Determination. a) Identification of terminal bases. The 5'-terminal base (green circle) and the 3'-terminal deoxyribose (red circle) exhibit only intranucleotide NOE's (as indicated by no other signals along the dashed lines). A 3'-terminal cytosine will have a second NOE signal corresponding to the connectivity between the adjacent C-6H, C-5H protons. This signal tends to be stronger than other signals and should not be confused as a 1'H, C-6H NOE. b) NOE links (horizontal lines) between the intranucleotide

connectivity $d_i(5CH_3 \rightarrow 6H)$ and the sequential connectivity $d_s(6H/8H \rightarrow 5CH_3)$. c) Partial sequences being matched to the DNA sequence. The remaining signals can then be identified more easily. 67

Figure 4.6: ^{1}H NMR Spectrum of 5'-CAT ATG-3'. Protons have been assigned as described above 68

Figure 4.7: Terminal Bases for 5'-CAT ATG-3'. The 5'-terminal cytosine base (green circle) and the 3'-terminal guanine deoxyribose (red circle). Note that there are no additional cross peaks along the dashed lines (the large cross peak on the vertical line (blue circle) is from the connectivity between C-6H, C-5H protons 69

Figure 4.8: S1A-8H→CH_3 – CH_3→S6T-6H and S2A-8H→CH_3 – CH_3→S7T-6H Links. Links are indicated by horizontal lines. 70

Figure 4.9: NOESY Assignment Link Pathway for 5'-CAT ATG-3' with $d_i(6H/8H \rightarrow 1'H)$ and $d_s(1'H \rightarrow 6H/8H)$ Connectivities. 71

Figure 4.10: NOESY Assignment Link Pathway for 5'-CAT ATG-3' with $d_i(6H/8H \rightarrow 2'H)$/ $d_i(6H/8H \rightarrow 2''H)$ and $d_s(2'H \rightarrow 6H/8H)$/$d_s(2H'' \rightarrow 6H/8H)$ Connectivities. Solid horizontal lines indicate 2'H i-s links and horizontal dashed lines represent 2"H i-s links (as seen side on). 72

Figure 4.11: Combined NOESY-COSY Assignment Link Pathway for 5'-CAT ATG-3' with $d_i(6H/8H \rightarrow 2'H)$ and $d_s(2''H \rightarrow 6H/8H)$ Connectivities. The signals forming the COSY spectrum are each comprised of two symmetrical cross-peaks (one positive, blue, and one negative, red), each along the one of the diagonals of the spectra. These cross peaks are a result of magnetisation transfer between two coupled protons. Where these diagonals cross is the centre of the overall signal. Note: some of these cross-peaks overlap or are not fully visible at this resolution. 73

Figure 4.12: Peak Broadening and Shifting. An example of the broadening of H1-NMR peaks during the melting process of 5'-CAT ATG-3'. The C1-6H signal starts as a sharp peak in the duplex form at 277.2 K before it broadens until almost flat at 286.2 K. It then proceeds to sharpen again in the final monomer form at 304.2 K. It can also be seen that the T3-6H and T5-6H peaks have crossed over each other. 75

Figure 4.13: Residue Melting Curves for the DNA Hexamer. 0.1 MPa – blue, 100 MPa – red, 150 MPa – green, 200 MPa – purple, 240 MPa – light blue. Melting curves are presented as the chemical shift (/ppm) as a function of temperature (/K). 76

Figure 4.14: An Incorrectly Fitted Curve. The melting curve (blue) of T3-6H has been fitted with equation (21) (red). The gradient of the fitted line below 280 K does not follow the trend of the line 78

Figure 4.15: Improved Fit. The melting curve (blue) of T3-6H has been fitted with equation (23) (red). The gradient of the fitted line below 280 K now has a more suitable gradient. 78

Figure 4.16: Melting Point versus Pressure for 5'-CAT ATG-3'. The curves for each base hydrogen are displayed along with the average melting point (bold black line). 80

Figure 4.17: CD Spectra During DNA Melting. The intensities at 208, 270 and 290 nm can be used to determine the melting point of the DNA. 81

Figure 4.18: CD Melting Curves. Melting curves obtained via CD spectroscopy at wavelengths of 208, 250, 270 and 290 nm. The melting point can be determined by the temperature corresponding to a normalised intensity of 0.5. 81

Figure 5.1: ^{1}H NMR Spectrum of Sequence 123 (5'-CAA GTC GAC TTG-3'). Protons have been assigned as described in the text. 92

Figure 5.2: Identification of Terminal Bases. The 5'-terminal cytosine base (green circle) and the 3'-terminal guanine deoxyribose (red circle). For the three cytosine signals (green dashed lines) it can be seen that there are two which have three vertical links while S7C-6H only has two (the third signal for S14C-6H is located at about 5.2 ppm and is not visible in this image). This is a C-6H→C-5H signal and is not relevant to the assignment of this sequence. After excluding the C-6H→C-5H signal, S7C-6H has only one link and is therefore the 5'-terminal base. Similarly for guanine, it can be seen that there is no horizontal link from the intranucleotide NOE on S4G-8H (red dashed line), identifying it as the 3'-terminal base. 93

Figure 5.3: S10G-8H→CH_3 – CH_3→S15T-6H, S14C-6H→CH_3 – CH_3→S11T-6H and S11T-6H→CH_3 – CH_3→S13T-6H Links. Links are indicated by solid horizontal lines. 93

Figure 5.4: NOESY Assignment Link Pathway for Sequence 123 (5'-CAA GTC GAC TTG-3') with d_i(6H/8H→1'H) and d_s(1'H→6H/8H) Connectivities. The link pathway can be followed from the dark red X along the lines in "rainbow" order through to the final black line. 95

Figure 5.5: NOESY Assignment Link Pathway for Sequence 123 (5'-CAA GTC GAC TTG-3') with d_i(6H/8h→2'H)/ d_i(6H/8H→2"H) and d_s(2H'→6H/8H)/d_s(2"H→6H/8H) Connectivities. Solid horizontal lines indicate 2' i-s links and horizontal dashed lines represent 2" i-s links (as seen side on). The link pathway can be followed from the dark red X along the lines in "rainbow" order through to the final black line. 96

Figure 5.6: Combined NOESY-COSY Assignment Link Pathway for Sequence 123 (5'-CAA GTC GAC TTG-3') with d_i(6H/8H→2H') and d_s(2"H→6H/8H) Connectivities. The signals forming the COSY spectrum are each comprised of two symmetrical cross-peaks (one positive, blue, and one negative, red), each along the one of the diagonals of the spectra. These cross peaks are a result of magnetisation transfer between two coupled protons. Where these diagonals cross is the centre of the overall signal. Note: some of these cross-peaks overlap or are not fully visible at this resolution. The link pathway can be followed from the dark red X along the lines in "rainbow" order through to the final black line. 97

Figure 5.7: Examples of Melting Curves Where Data Gathering Proved Difficult. a) An incomplete curve where the signal has been followed from either end of the melting process

to establish an almost full curve. b) A curve where the missing signals in the melting region have left insufficient data for accurate fitting of the melting curve..... 99

Figure 5.8: Residue Melting Curves for Sequence 121. 0.1 MPa – blue, 50 MPa – red, 100 MPa – green, 150 MPa – light blue, purple. Melting curves are presented as the chemical shift (/ppm) as a function of temperature (/K). 100

Figure 5.9: Residue Melting Curves for Sequence 122. 0.1 MPa – blue, 50 MPa – red, 100 MPa – green, 150 MPa – orange, light blue, purple, 200 MPa – light purple. Melting curves are presented as the chemical shift (/ppm) as a function of temperature (/K). 101

Figure 5.10: Residue Melting Curves for Sequence 123. 0.1 MPa – blue, 50 MPa – red, 100 MPa – green, 150 MPa – orange, light blue, purple, 200 MPa – light purple. Melting curves are presented as the chemical shift (/ppm) as a function of temperature (/K). 102

Figure 5.11: Residue Melting Curves for Sequence 124. 0.1 MPa – blue, 50 MPa – red, 100 MPa – green, 150 MPa – light blue, purple. Melting curves are presented as the chemical shift (/ppm) as a function of temperature (/K). 103

Figure 5.12: Residue Melting Curves for Sequence 125. 0.1 MPa – blue, 50 MPa – red, 100 MPa – green, 150 MPa – light blue, purple, 200 MPa – light purple. Melting curves are presented as the chemical shift (/ppm) as a function of temperature (/K). 104

Figure 5.13: Melting Point versus Pressure for Sequence 121 (5'-CAT TTA TAA ATG-3'). Plots for each residue are displayed along with the average (thick black line). Plots where melting points could only be approximately estimated are shown as dashed lines. 106

Figure 5.14: Melting Point versus Pressure for Sequence 122 (5'-CAT TCT AGA ATG-3'). Plots for each residue are displayed along with the average (thick black line). Plots where melting points could only be approximately estimated are shown as dashed lines. 107

Figure 5.15: Melting Point versus Pressure for Sequence 123 (5'-CAA GTC GAC TTG-3'). Plots for each residue are displayed along with the average (thick black line). Plots where melting points could only be approximately estimated are shown as dashed lines. 108

Figure 5.16: Melting Point versus Pressure for Sequence 124 (5'-CAG GTC GAC CTG-3'). Plots for each residue are displayed along with the average (thick black line). Plots where melting points could only be approximately estimated are shown as dashed lines. 109

Figure 5.17: Melting Point versus Pressure for Sequence 125 (5'-CAC CCG CGG GTG-3'). Plots for each residue are displayed along with the average (thick black line). Plots where melting points could only be approximately estimated are shown as dashed lines. 110

Figure 5.18: CD Spectra of Sequence 122 During DNA Melting. 111

Figure 5.19: CD Melting Curves. a) Melting curves obtained via CD spectroscopy at wavelengths of 208, 245 and 277 nm for sequence 122. The melting point can be determined by the temperature corresponding to the average normalised intensity of 0.5. Also presented are

sequences b) 121, c) 123, d) 124 and e) 125. CD melting spectra and larger melting curve images are available in Appendix 3. 112

Figure 5.20: Melting Points for Each Residue of Sequence 121 (5′-CAT TTA TAA ATG-3′) Plotted as a Function of Sequence Position. Those melting points which could not be accurately determined are represented by unfilled markers. 115

Figure 5.21: Melting Points for Each Base of Sequence 122 (5′-CAT TCT AGA ATG-3′) Plotted as a Function of Sequence Position. Those melting points which could not be accurately determined are represented by unfilled markers. 117

Figure 5.22: The Partial Molar Isothermal Compressibilities at Infinite Dilution $K°_{T,2}$ versus Pressure for Adenosine and Thymidine at 298.2 K [81, 82]. 119

Figure 5.23: The Partial Molar Isothermal Compressibilities at Infinite Dilution $K°_{T,2}$ versus Pressure for Cytidine and Guanosine at 298.2 K [82, 83]. 119

Figure 5.24: Melting Points for Each Base of Sequence 123 (5′-CAA GTC GAC TTG-3′) Plotted as a Function of Sequence Position. Those melting points which could not be accurately determined are represented by unfilled markers. 121

Figure 5.25: Melting Points for Each Base of Sequence 124 (5′-CAG GTC GAC CTG-3′) Plotted as a Function of Sequence Position. Those melting points which could not be accurately determined are represented by unfilled markers. 123

Figure 5.26: Melting Points for Each Base of Sequence 125 (5′-CAC CCG CGG GTG-3′) Plotted as a Function of Sequence Position. Those melting points which could not be accurately determined are represented by unfilled markers. 125

Figure 6.1: Triplex Base Pairing and Folding. a) The T-A*T triplex base pairing. b) The C-G*C^+ base pairing. These triplex pairings are formed when the central A or G base from both a Watson-Crick base pairing (T-A of C-G respectively) combined with a Hoogsteen base pairing (A*T or C*G respectively). c) The folded triplex structure. A section of the sequence has dissociated from its original duplex pairings and folded back along the chain forming Hoogsteen pairings (red lines) along the Watson-Crick (blue lines) paired duplex. 132

Figure 6.2: G-Quadruplex Base Pairing and Structures. a) The tetrad formed by four guanine residues. b) A G-quadruplex structure formed from four locally continuous oligonucleotides. c) A G-quadruplex structure formed from two separate folded oligonucleotides (one of many different configurations). d) A G-quadruplex structure formed from a single folded oligonucleotide containing three lateral loops. e) A G-quadruplex structure formed from two lateral loops and a diagonal loop, f) A G-quadruplex structure formed from two lateral loops and a propeller loop (the loop around the side of the structure). g) A G-quadruplex structure formed from three propeller loops. The topologies of these structures are only a sample of all the different possibilities. 133

Figure 6.3: i-Motif Base Pairing and Structure. a) the C-C^+ base pairing. b) An i-motif structure formed from four separate oligonucleotides. c) An i-motif structure formed from two separate folded oligonucleotides. d) An i-motif structure formed from a single folded oligonucleotide. 133

Figure 6.4: i-Motif ^{1}H NMR Spectrum. The types of base for each proton signal have been identified. 139

Figure 6.5: i-Motif Spectrum at 285.2 K and pH 6.4. The H^+ proton signal has been magnified to show the fine detail. These signals disappear during the melting of the i-motif. 140

Figure 6.6: Melting Curve for pH 6.4 at 0.1 MPa. The intensity of the H^+ proton peaks relative to the NMR standard has been plotted as a function of temperature along with the fitted melting curve from equation (23). 141

Figure 6.7: i-Motif Melting Points as a Function of Temperature in Phosphate Buffer. Melting points obtained via fitting at pressures of 0.1, 100 and 200 MPa are given. Values of pH are uncorrected for pressure-induced shifts. Lines between data points are included as a visual reference only. 142

Figure 6.8: i-Motif Melting Points as a Function of Temperature in Citrate Buffer. Melting points obtained via fitting at pressures of 0.1, 50, 100, 150 and 200 MPa are given. Values of pH are uncorrected for pressure-induced shifts. Lines between data points are included as a visual reference only. 142

Figure 6.9: CD Melting Curves for the i-Motif at pH 4.6 and 0.1 MPa. Melting curves obtained via CD spectroscopy at wavelengths of 222 nm, 255 nm and 290 nm. The melting point can be determined by the temperature corresponding to the average normalised intensity of 0.5.. 143

Figure 6.10: Melting Point versus pH for NMR and CD Spectroscopy in Citrate Buffer at 0.1 MPa. Lines are for visual reference only. 144

Figure 6.11: i-Motif Melting Points as a Function of Temperature in Phosphate Buffer with Corrected pH Values. Melting points obtained via fitting at pressures of 0.1, 100 and 200 MPa are given with pH values at high pressures corrected for pressure-induced pH shifts. Lines are for visual reference only. 145

Figure 6.12: Decrease in Melting Point of the i-Motif at 200 MPa. The decrease in the melting point of the i-motif at 200 MPa has been plotted as a function of corrected pH. 146

Figure 6.13: i-Motif Melting Points as a Function of Temperature in Citrate Buffer with Corrected pH Values. Melting points obtained via fitting at pressures of 0.1, 50, 100, 150 and 200 MPa are given with pH values at high pressures corrected for pressure-induced pH shifts. Lines are for visual reference only. 147

Figure 6.14: i- Motif Melting Point versus Pressure at pH 5.85. The data have had no corrections for pressure-induced pH shifts. The corrected data have been obtained by

calculating the decrease in melting point at each pressure compared to the 0.1 MPa value at the same pH. The melting points have then been plotted relative to the pH 5.85 0.1 MPa reading. 148

Figure 7.1: SBS Sequencing Image. Each of the differently coloured dots in this image corresponds to a given DNA base. Computer software will examine each sequencing image to construct the sequence for all the DNA strands on this surface. Image obtained from Yale Centre for Genome Analysis. 161

Figure 7.2: Illumina Preparation and Sequencing. Genomic DNA is first sheared to lengths of 200-300 bp. Adapters are added to the ends of the DNA fragments and then flowed through a flow cell where they bind to designated sites. Bridge amplification is then used to generate DNA clusters in preparation for sequencing by synthesis. Image adapted from Bite Sized Bio, http://bitesizebio.com/13546/ sequencing-by-synthesis-explaining-the-illumina-sequencing-technology/. 163

Figure 7.3: Gel Electrophoresis Image of the Longest Incubation Time for Each Temperature and Pressure. Lane 1, 1 kb+ ladder. Lanes 2-3, SS DNA at 0.1 MPa and 150 MPa respectively at 283.2 K. Lanes 4-7, SS DNA at 0.1 MPa, 50 MPa, 100 MPa and 150 MPa respectively at 313.2 K. Lanes 8-9, SS DNA at 0.1 MPa and 150 MPa respectively at 333.2 K. Lanes 10-13, RFII DNA at 0.1 MPa, 50 MPa, 100 MPa and 150 MPa respectively at 333.2 K. Lanes 14-15, RFI DNA at 0.1 MPa and 150 MPa respectively at 333.2 K. 165

Figure 7.4: Post-Amplification Gel Electrophoresis Image. Lanes 1-9, 0.1 MPa, 283.2 K, SS DNA. Lanes 10-18, 150 MPa, 283.2 K, SS DNA. Lanes 19-22, 0.1 MPa, 313.2 K, SS DNA. Lanes 23-26, 50 MPa, 313.2 K, SS DNA. Lanes 19-22, 0.1 MPa, 313.2 K, SS DNA. 167

Figure 7.5: Gel Electrophoresis Image for Samples Which Failed to Amplify. Lanes correspond to samples 5-9,10, 19-22 and 30 from Figure 7.4 respectively. 167

Tables

Table 2.1: Buffer Reaction Volumes (ΔV^0). This table shows the pK_a and reaction volumes [51] of buffers relevant to this research. 29

Table 2.2: Buffer pH Corrections. Corrected atmospheric pH values (pH_0) for pH 7.0 phosphate buffer for commonly used pressure levels at 298.2 K. 30

Table 3.1: Table of Stock Solutions for Cytosine/Cytidine Experiments. Each entry shows concentration, mass of compound used, volume made (in milli-Q water) and supplier. The pH value given with the NaCl solutions indicate the pH at which the solution is to be used. 37

Table 3.2: Buffer pH Corrections. This table displays the buffer solutions used for these studies. The first two columns show the intended pressure conditions and the pH required. The third

column shows the resulting pH shift and the fourth column shows the pH value at 0.1 MPa required to achieve the correct pH under pressure. 38

Table 3.3: Salt Concentrations. This table shows the salt concentrations required to achieve an ionic strength of 0.2 M for the solutions of different pH. The concentration of the initial solutions before dilution with the cytosine/cytidine solution is also shown. 40

Table 3.4: Conditions for Incubation of Cytosine. This table displays the calculated k_{obs} values for the hydrolysis of cytosine for each pressure and pH condition at 373.2 K. The half-life ($t_{1/2}$) in days is also shown below each k_{obs} value. 46

Table 3.5: Conditions for Incubation of Cytidine. This table displays the calculated k_{obs} values for the hydrolysis of cytidine at 0.1 MPa and 150 MPa at pH 7. The half-life ($t_{1/2}$) in days is also shown below each k_{obs} value. 49

Table 4.1: Table of Stock Solutions for Hexamer Melting Experiments. Each entry shows concentration, mass of compound used, volume made (in milli-Q water) and supplier. 60

Table 4.2: Sequencing Assignment for 5'-CAT ATG-3'. 70

Table 5.1: Dodecamer Sequences. Sequences of dodecamers used along with GC content, concentrations used and reference number. 88

Table 5.2: Table of Stock Solutions for Dodecamer Melting Experiments. Each entry shows concentration, mass of compound used, volume made (in milli-Q water) and supplier. 89

Table 5.3: Assignment of Sequence 123. 94

Table 5.4: CD Melting Results of Dodecamers. The CD melting temperature results are presented here alongside the results obtained from NMR spectroscopy at 0.1 MPa. 113

Table 5.5: Visual Representation of T_m Drops. T_m drops appear between junctions between consecutive sequences of purines (purple) and pyrimidines (blue). A drop in the melting point of a base is represented by a coloured square below it. Yellow represents a drop of 1-2 K, orange represents a drop of 3-4 K and red represents a drop of > 4 K. A * indicates a melting point which was estimated but agrees with its complementary melting point. A ** indicate an estimated melting point which does not match that of their complementary base. In these instances they have instead been given the T_m drop of the complementary base. 126

Table 6.1: Table of Stock Solutions for i-Motif Melting Experiments. Each entry shows concentration, mass of compound used, volume made (in milli-Q water) and supplier. 137

Table 6.2: Buffer pH Values for NMR Spectroscopy Melting Experiments. The pH values for each citrate and phosphate buffer used are shown here. 137

Table 7.1: ΦX174 Incubation Conditions. This table outlines the temperature and pressure conditions each form of ΦX174 was subjected to, along with the maximum period of

incubation time in days (brackets) and number of time points taken during this time (square brackets). 155

Table 7.2: Incubation Times and Temperatures. Also given are the corresponding rate constants, number of predicted hydrolysis events per cytosine and the average number of predicted cytosine hydrolysis events in 50,000 chains at 0.1 MPa. The last column is for the double-stranded sequence which were incubated at higher temperatures and having a rate constant 140 times lower than its single-stranded counterpart..... 156

Table 7.3: Table of Buffer Stock Solutions for Chemical Stability Experiments. Each entry shows concentration, mass of compound used, volume made (in milli-Q water) and supplier..... 156

Table 7.4: Buffer pH Corrections. This table displays the buffer solutions used for these studies. The first two columns show the intended pressure conditions and the pH required. The third column shows the resulting pH shift and the fourth column shows the pH value at 0.1 MPa required to achieve the correct pH under pressure..... 157

Table 7.5: PCR Mix. This table shows each component added to the PCR mixture. Each component's initial concentration, volume added and final concentration are given. The final solution was made up to 40 µL using milli-Q water. 159

Table 7.6: PCR Protocol. Temperatures, times and cycles used during PCR amplification of the incubated DNA samples. 160

List of Nomenclature and Abbreviations

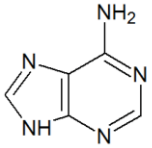
Nomenclature

The building blocks of DNA/ RNA

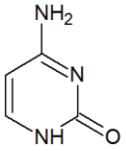
- Nucleobases

Common Bases

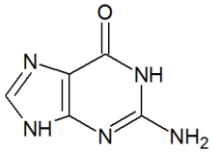
- Adenine



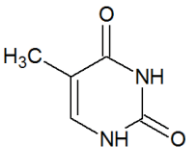
- Cytosine



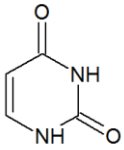
- Guanine



- Thymine

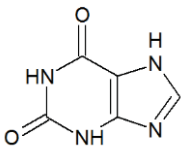


- Uracil

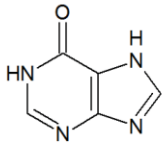


Other Bases

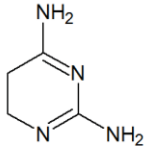
- Xanthine



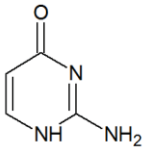
- Hypoxanthine



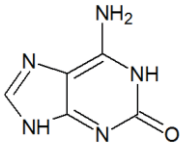
- Diaminopyrimidine



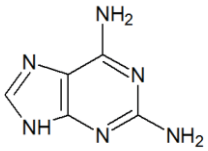
- Isocytosine



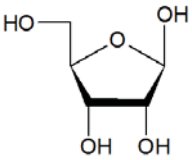
- Isoguanine



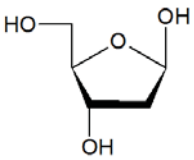
- Diaminopurine



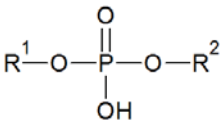
- Sugars
  - Ribose



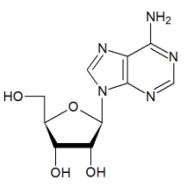
  - 2-Deoxyribose



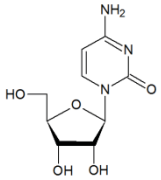
- Phosphodiester



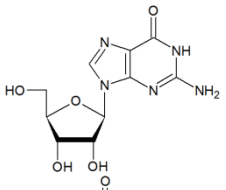
- Nucleosides
  - Adenosine



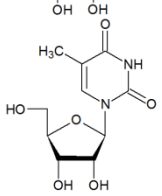
- Cytidine



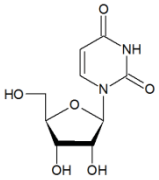
- Guanosine



- Thymidine



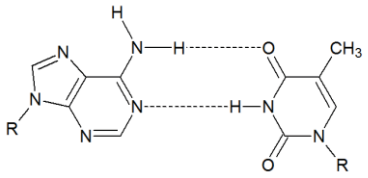
- Uridine



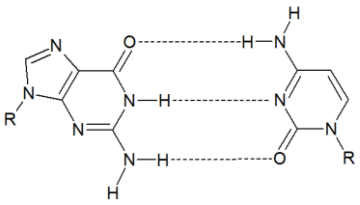
- Base pairing

  Watson-Crick

  - A-T

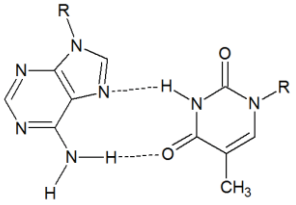


  - G-C

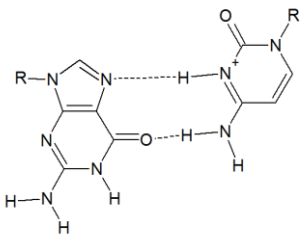


  Hoogsteen

  - A*T

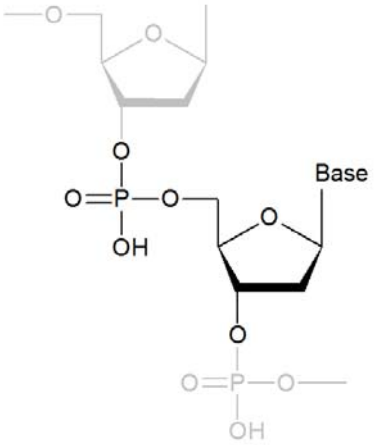


  - G*C⁺

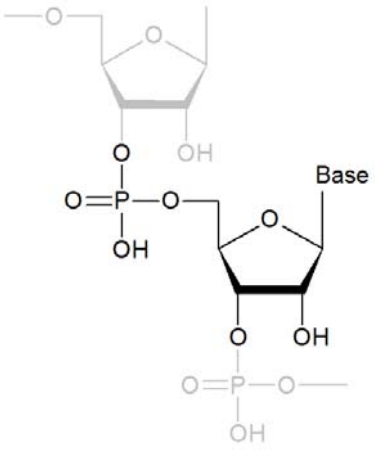


Nucleic Acids

- DNA - Deoxyribonucleic acid. Each DNA nucleotide is comprised of a nucleobase, a phosphate and a 2-deoxyribose sugar. The phosphate and sugar form the backbone of DNA. The nucleobases are attached to the anomeric carbon of the sugar unit.
    - DNA nucleotide

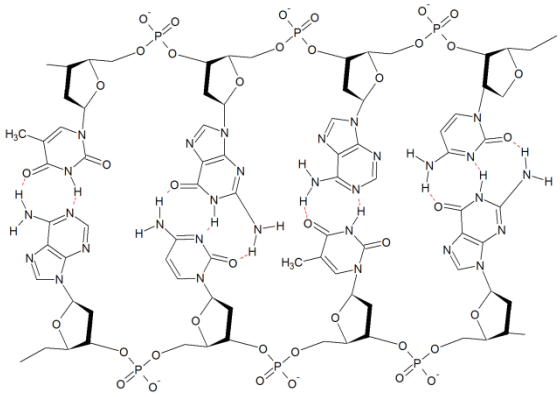


- RNA - Ribonucleic acid. Each RNA nucleotide is comprised of a nucleobase, a phosphate and a ribose sugar. The phosphate and sugar form the backbone of RNA. The nucleobases are attached to the anomeric carbon of the sugar unit.
    - RNA nucleotide



- Nucleic acid structure. A series of nucleotides joined to form a oligonucleotide chain. Watson-Crick base pairing between two

complementary oligonucleotides under standard biological conditions results in the DNA structure as show below. Three dimensionally this is a right-handed double helix with approximately 10.5 bases per turn.

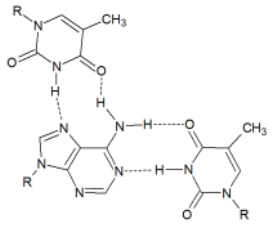


- Other DNA structures

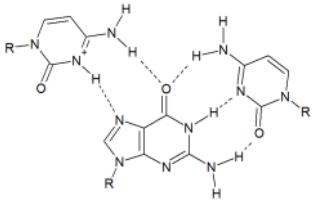
Triplexes. A triplex is formed when a C/T rich segment of a DNA sequence folds back along a A/G rich segment. This results in a triplex comprised of a pair of strands bonded by Watson-Crick pairings with a third strand bound via Hoogsteen pairings to the A/G rich strand involved in the Watson-Crick pairing.

  - Triplex pairing

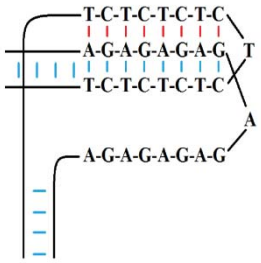
T-A*T



C-G*C+

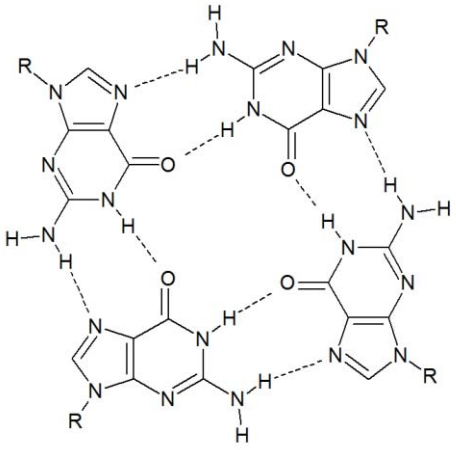


- Triplex Structure

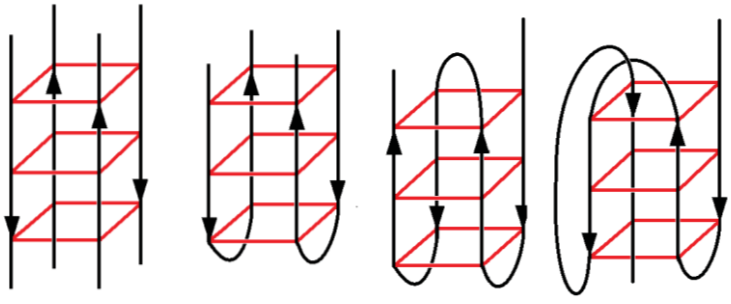


G-Quadruplexes. G-quadruplexes are a tetrad structure formed by four, G-rich sequences. Four guanine bases undergo Hoogsteen type bonding to form a G tetrad. These tetrads stack to form a quadruplex with multiple topologies.

- G-quadruplex tetrad

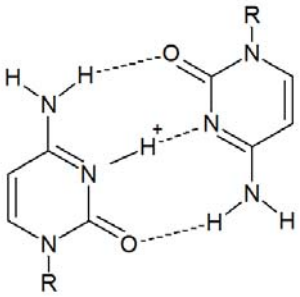


- G-quadruplex Topologies

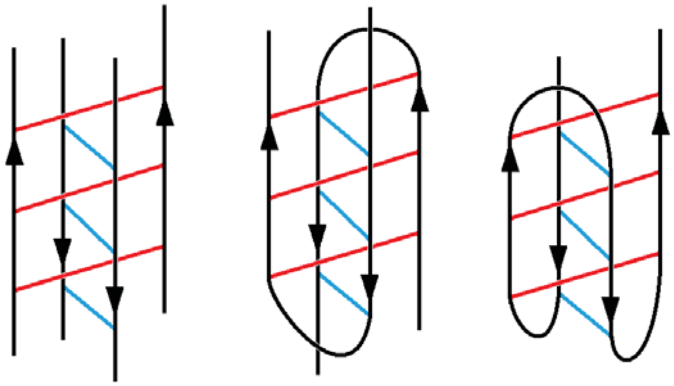


i-Motifs. i-Motifs are a tetrad structure formed by four, C-rich, DNA sequences. The structure is formed from two sets of antiparallel stranded duplexes connected by C-C^+ Hoogsteen pairings. These pairings are intercalated between the two duplexes resulting in the four stranded structures.

- i-Motif pairing



- i-Motif structures



Abbreviations

CD spectroscopy	- Circular Dichroism spectroscopy
COSY	-Correlation Spectroscopy
dNTPs	- deoxyribonucleotide Triphosphates
HiP	- High Pressure Equipment Company
NMR spectroscopy	- Nuclear Magnetic Resonance spectroscopy
NOESY	-Nuclear Overhauser Effect Spectroscopy
RF	- Radio Frequency
SBS	- Sequencing By Synthesis
Sequence 121	- DNA sequence 5′-CAT TTA TAA ATG-3′. 16.7 % G-C
Sequence 122	- DNA sequence 5′-CAT TCT AGA ATG-3′. 33.3 % G-C
Sequence 123	- DNA sequence 5′-CAA GTC GAC TTG-3′. 50 % G-C
Sequence 124	- DNA sequence 5′-CAG GTC GAC CTG-3′. 66.7 % G-C
Sequence 125	- DNA sequence 5′-CAC CCG CGG GTG-3′. 83.3 % G-C
T_m	- Melting temperature
TMSP-d	- 3-(trimethylsilyl)-2,2',3,3'-tetradeuteropropionic acid (NMR reference)